

# nature

THE INTERNATIONAL JOURNAL OF SCIENCE

Hawaiian gull  
costal records  
North Pacific  
nitrogen levels  
Following  
Little Ice Age

1000

## SEA CHANGE

THIS YEAR'S  
HOT TICKETS

Climate change and more

1000



LEARNING  
AGGRESSION

How the brain's learning system

1000

NEANDERTHAL  
WOMAN

Aggression in prehistoric

1000

SCIENCE

1000

1000

# THIS WEEK

## EDITORIALS

**WORLD VIEW** A trip to Brazil shows Colin Macilwain the future **p.7**

**SMOKERS** Europe opts for lax regulation of electronic cigarettes **p.8**



**LEUKAEMIA** Disease-genetics pioneer Janet Rowley dies aged 88 **p.8**

## The greater good

*Governments, funding agencies and universities must all do their bit to ensure that research is appropriately assessed and rewarded.*

Do researchers need extra incentives? The idea should seem strange to any young scientist — and a good many older ones too. Think of the theoretical physicist who will happily sit down at the weekend, not to read the newspapers but to play around with equations. Think of the cell biologist willing to put up with the burden of running and re-running painstaking experiments because it's so difficult to make the set-up deliver — but whose instincts suggest that there is a nugget of insight at the end.

Now think of the head of an institute of, say, plant science. He or she will have no shortage of talented researchers wanting to understand fundamentals such as the precise mechanisms of influence of plant hormones at various stages of plant development. But what criteria will be used to assure the university or management board that the head has delivered? And what incentives will encourage the broadening of research in directions that scientific insight left to its own devices might not prioritize but that might, nevertheless, serve humanity well?

Young scientists and lab heads alike live in a world in which the social contract for science is changing all around them. This is happening in ways they can influence, if they are both lucky and astute, and from which they could and should benefit. Astute people, of course, often make their own good luck — finding themselves in the right place at the right time by being alert to the way the world is moving and engaging more broadly with interests around their disciplines than less adventurous academics might.

### CHALLENGING CRITERIA

National governments — important drivers of the social contract — have the power to do more than steer the direction of science through broad funding priorities. They can, through their funding agencies, seek to ensure at least two other outcomes: an appropriate assessment and reward of research achievement, and an appropriate degree of trust in the robustness of that achievement.

For the former, there is plenty of action, but plenty of debate too. The head of the plant institute will know that it will be all too easy for assessors to focus excessively on the number of papers published in high-impact journals. (*Nature*, proud of its own high- and low-cited papers, has long challenged that inclination.) It is ever more important in researcher assessments to recognize the work that focuses on key societal challenges. Examples across the spectrum could include exploring how established techniques can enhance plant resistance to disease; work on climate-change adaptation; studies to enhance access to fresh water; and the development of psychological treatments for post-traumatic stress. Such work may at times be scientifically incremental, but has every bit as much claim on recognition.

Some major universities are recognizing the importance of such challenges by establishing their own programmes that may include the natural and social sciences and humanities (see page 7). For example, the National University of Singapore has a cluster of research groups

working on the future of high-density cities; the University of Cambridge, UK, has several departments collaborating in public health; and Monash University in Melbourne, Australia, has established a programme on aspects of fresh water.

Such collaborations are not easy to make effective, and it is therefore doubly important that 'the system' finds new ways to recognize and reward their outputs. In this respect, an interesting case to follow in 2014 is the United Kingdom's Higher Education Funding Council for England (HEFCE) and its dependent institutions. Its national research assessments have evolved over the years in ways that other agencies around the world have examined — although few, if any, have imitated the extreme extent to which the outcomes directly influence subsequent funding. But this year, for the first time, we shall see what the HEFCE has made of the thousands of statements of research impacts submitted last year by universities.

Researchers have often expressed alarm at the perceived tendency in such exercises to move from one extreme focus of assessment — journal impact factors — to another, in particular the anticipation of contributions to economic growth. This year will show whether the HEFCE can demonstrate the appropriate degree of breadth, nuance and critical assessment of such statements, while giving due recognition to the socially valuable work advocated above.

And what about research robustness — in other words, the trust that taxpayers can have that the appropriate standards of technical integrity, aka professionalism, are being followed in laboratories? Journals, not least *Nature*, are recognizing that they and their referees have a part to play in ensuring better standards of quantitative analysis, and of data and protocol transparency. Universities have a key role too — in ensuring a greater degree of researcher training and of lab stewardship in the quality of outputs than is often happening. That is a tough challenge, because vice-chancellors have so little power over how their academics behave, and those academics are all too often engaged in a rat race for funds.

But 2014 should be a year in which funding agencies make clear their intentions in promoting rigorous lab standards, and there should be a concomitant pressure on universities and institutes to demonstrate quality assurance of lab practices and culture.

What is essential is that the motivation of young scientists to make a difference with their research is more broadly encouraged. They need strong mentoring and exemplars in doing a robust job and in contributing to the trust in their research community. Those who want to follow strong creative imaginations in discovering how the world works should be given full rein. So, too, should those more interested in using their creativity directly to make the world a better place. Those who have the luck to be able to do both and to be recognized for both will be in a sweet spot indeed. ■

**“National governments have the power to do more than steer the direction of science.”**





## Emerging powers need a more-inclusive science

*Fast-growing economies can learn from the West's mistakes and couple social and 'hard' sciences to address their own societal needs, says Colin Macilwain.*

Rio de Janeiro's peerless Copacabana beach has not changed much in the 15 years since I last visited, but there has been one innovation: all along the promenade, sturdy, open-air gym facilities invite locals and tourists alike to indulge in a little anaerobic exercise.

This free equipment seems an obvious approach to improving public health. But tell that to inhabitants of the Bronx in New York or of London's East End, where the most visible signs of health science are the nearby glass towers in Washington Heights or Whitechapel, in which biologists develop drugs that largely benefit the well-to-do.

Rio's outdoor gyms reflect the work of researchers such as epidemiologist Pedro Hallal at the Federal University of Pelotas in southern Brazil, who is part of an influential movement to better understand the links between mothers' health, early-childhood exercise and lifelong health outcomes. This is the sort of societal research that the developing world needs as it expands its scientific influence.

"Let's forge this connection between the social science and the hard sciences," Michel Temer, vice-president of Brazil, told the 6th World Science Forum in Rio de Janeiro — the reason for my visit in November. The point was forcefully reiterated by Linxiu Zhang of the Chinese Academy of Sciences, and many other speakers.

Make no mistake: the geographical balance of power in global science is shifting. China has surpassed the United States as the world's largest PhD factory (see *Nature* **472**, 276–279; 2011) and about now, according to a 2011 report by Britain's Royal Society, it is scheduled to surpass the US volume of scientific literature in research journals. Brazil awarded 14,000 PhDs last year.

The shift is accompanied by real political determination from the emerging powers to couple the social sciences with 'hard' science and engineering to address society's needs. For their own pressing political reasons, the leaders of Brazil, China and other fast-growing economies need answers to mounting societal problems — water, food, health, energy and climate change, for example. That is not the case in the United States or Europe, where leaders' priorities are short-term and financial, and science is arranged to suit various stakeholders — notably firms that supply drugs and military equipment — as well as the needs of scientists themselves and their universities.

There are well-charted historical reasons for the West's narrow view of what constitutes science. Around 1900, scientists of the Royal Society of London distanced themselves from colleagues in the humanities (leading to the formation of the British Academy), and the US National Academy of Sciences followed the same path.

The outcome has been subtler in mainland

Europe. The German word for science, *Wissenschaft*, acknowledges a wider body of knowledge than just the natural sciences, for example; and the former president of the prestigious European Research Council, Helga Nowotny, is a sociologist.

Yet the question of fair treatment for the social sciences is dogging the new European Union (EU) research programme Horizon 2020, the largest in the world outside the United States. Social scientists feel that they have been locked out of the drafting of the Horizon 2020 work programmes. At a 26 November meeting in Brussels on 'smart cities', for example, speakers castigated the planned programme for concentrating on technology-led pilots, even though the real roadblock is how people use the technologies we already have.

These are not abstract, philosophical questions: quantitative behavioural research could readily fill knowledge gaps and design processes that would enable people to better manage their energy use, for example. But it does not happen because EU research programmes are also designed around the needs of stakeholders: in this case, device manufacturers, power companies and university scientists and engineers who know the ropes from previous programmes.

Another closely associated issue raised at the Rio meeting is the fact that global science still has a huge problem with research 'silos', in which researchers are obliged to operate within insular, sometimes archaic disciplines. This was broached by physicist Luiz Davidovich, a director of the

Brazilian Academy of Sciences in Rio, who called for the "reformulation of the university, towards interaction between disciplines". But the West's funding agencies and universities — as well as its publishing industry — are all set up in ways that have persistently stymied such change. An opportunity surely exists for emerging scientific powers to do things differently as they grow, by building an interdisciplinary outlook into their structures.

The World Science Forum is just one instrument that is attempting to address such problems. In 2012, the Global Research Council was created at the instigation of Subra Suresh, then director of the US National Science Foundation, as a vehicle for the wider governance of science.

Existing worldwide organizations have limited influence, however. The new global agenda is more likely to be driven by the most powerful of the emerging powers: China, in particular, but also Brazil, India, South Korea and South Africa. That group of emerging nations has the opportunity, right now, to build a science that will serve not just the interests of national oligarchies, or of researchers themselves, but of society at large. ■

**Colin Macilwain** writes about science policy from Edinburgh, UK.  
e-mail: [cfmworldview@googlemail.com](mailto:cfmworldview@googlemail.com)

**MAKE NO MISTAKE:  
THE GEOGRAPHICAL  
BALANCE OF  
POWER  
IN GLOBAL SCIENCE IS  
SHIFTING.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/af8erx](http://go.nature.com/af8erx)

# SEVEN DAYS

The news in brief

## POLICY

### EU clinical trials

A long-running effort to reform the regulation of clinical trials in the European Union concluded on 20 December. A new law will replace the much-maligned Clinical Trials Directive, and will streamline and standardize applications for trials. The Clinical Trials Regulation, which includes compulsory preregistration of all trials and tougher informed-consent requirements, must be formally approved before it can take effect.

### Primate problems

The US Department of Agriculture has fined Harvard Medical School in Boston, Massachusetts, more than US\$24,000 for violating the Animal Welfare Act. Several of the 11 violations announced by the agency on 18 December occurred at Harvard's troubled New England Primate Research Center in Southborough, Massachusetts, which is slated to close by mid-2015. In late 2011, two primates became dehydrated as the result

of a malfunctioning water dispenser. One of the animals subsequently died.

### E-cigarette rules

European legislators have shied away from tough proposals to regulate electronic cigarettes as medical devices. According to legislation agreed on 18 December, e-cigarettes will be subject to the less-stringent controls that are applied to tobacco products, unless they are marketed with health claims. Formal approval of the agreement, which includes policy changes for tobacco products, is needed before the rules can come into force in 2014. See [go.nature.com/wobgkx](http://go.nature.com/wobgkx) for more.

## PEOPLE

### Geneticist dies

Geneticist Janet Rowley (pictured) of the University of Chicago in Illinois died on 17 December, aged 88. In the 1970s, Rowley identified a translocation — in which genetic material is swapped between chromosomes — in leukaemia cells. For her work on that and other translocations, Rowley was



one of three scientists to share a Lasker Award in 1998 for clinical medical research. From 2002 to 2009, she served on former US President George W. Bush's bioethics council, and was a vocal opponent of the Bush administration's restrictions on US embryonic stem-cell research.

### EPA fraudster

John Beale, a former official at the US Environmental Protection Agency, was sentenced on 18 December to 32 months in prison for stealing nearly US\$900,000 from the organization. Starting in 2000, Beale collected a salary and travel reimbursements while skipping a total of 2.5 years of work and falsely claiming to be working for the US Central

Intelligence Agency. Beale has agreed to pay the government nearly \$1.4 million in restitution and penalties.

### UCSF chancellor

Susan Desmond-Hellmann will step down as chancellor of the University of California, San Francisco (UCSF), in March to become chief executive of the Bill & Melinda Gates Foundation in Seattle, Washington. UCSF said on 17 December that it will appoint Sam Hawgood, dean of its medical school, as interim chancellor, pending approval.

## EVENTS

### Star surveyor

The European Space Agency's Gaia mission, designed to map the Milky Way in unprecedented detail, launched on 19 December from Kourou, French Guiana. The mission is tasked with charting a billion stars, with the aim of helping astronomers to better understand the origins of our Galaxy. See [go.nature.com/bhxyuq](http://go.nature.com/bhxyuq) for more.

## BUSINESS

### Pharma fees out

In a bid to mend its public image, pharmaceutical giant GlaxoSmithKline (GSK) will by 2016 start to phase out direct payments to physicians for attending medical conferences or giving promotional talks about GSK products, the firm said on 17 December. By early 2015, it also plans to scrap the use of prescription sales targets in determining pay for sales representatives. GSK currently faces investigation in China for allegedly bribing physicians and officials to boost its drug sales (see *Nature* 499, 385; 2013).

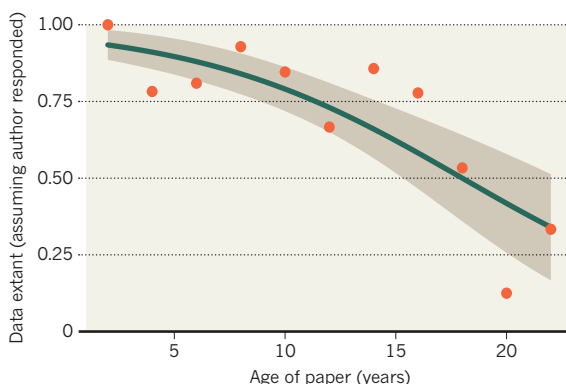
**NATURE.COM**  
For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

Raw data from research publications are vanishing rapidly, according to an analysis of 516 ecology papers published between 1991 and 2011 (T. H. Vines *et al.* *Curr. Biol.* <http://doi.org/qpm>; 2013). Data could be obtained for most of the 2011 papers, but availability fell by 17% for each previous year (see chart). As few as 20% of authors of papers from the early 1990s could provide data, as a result of the information being misplaced or stored on defunct technology. See [go.nature.com/jmosxn](http://go.nature.com/jmosxn) for more.

### MISSING DATA

As research articles age, the odds of their raw data being extant drop noticeably.





# NEWS IN FOCUS

**CLIMATE** Water will be scarce in warmer world, huge study says **p.10**

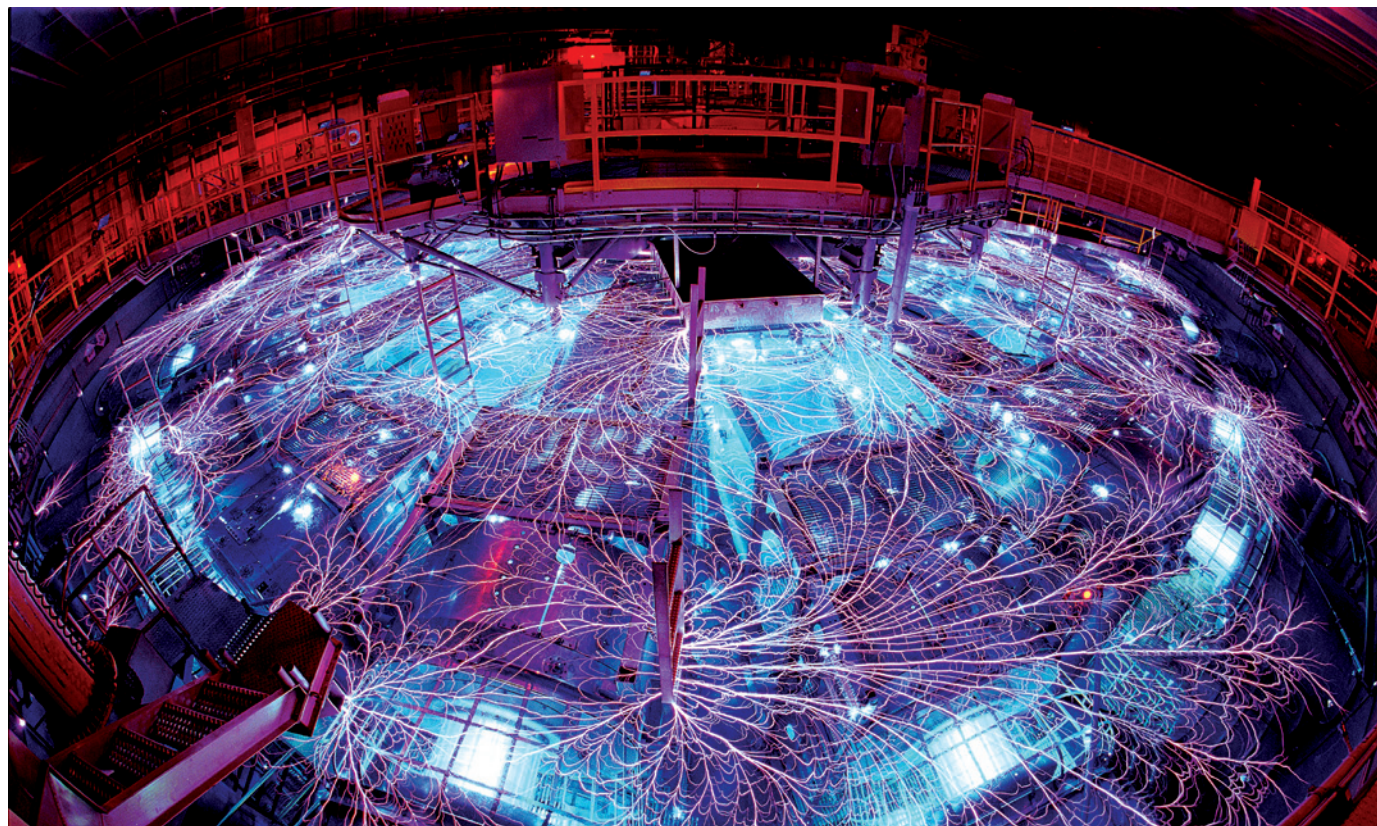
**PHYSICS** With \$50 million spent, X-ray source is now out in the cold **p.11**

**2014** Transgenic monkeys, comets and more will make the news this year **p.13**



**EPIGENETICS** A scientist's bid to track down the roots of aggression **p.14**

RANDY MONTVOYA/SANDIA NATL LAB.



The intense electrical discharge of the Z machine at Sandia National Laboratories in New Mexico is used in attempts to trigger nuclear fusion.

## PLASMA PHYSICS

# Triple-threat method sparks hope for fusion

*The secrets to its success are lasers, magnets and a big pinch.*

BY W. WAYT GIBBS

**T**he Z machine at Sandia National Laboratories in New Mexico discharges the most intense pulses of electrical current on Earth. Millions of amperes can be sent towards a metallic cylinder the size of a pencil eraser, inducing a magnetic field that creates a force — called a Z pinch — that crushes the cylinder in a fraction of a second.

Since 2012, scientists have used the Z pinch to implode cylinders filled with hydrogen isotopes in the hope of achieving the extreme temperatures and pressures needed for energy-generating nuclear fusion. Despite their efforts, they have never succeeded in reaching ignition — the point at which the energy gained from fusion is greater than the energy put in.

But after tacking on two more components, physicists think they are at last on the right path.

Researchers working on Sandia's Magnetized Liner Inertial Fusion (MagLIF) experiment added a secondary magnetic field to thermally insulate the hydrogen fuel, and a laser to pre-heat it (see 'Feeling the pinch'). In late November, they tested the system for the first time, using 16 million amperes of current, a 10-tesla magnetic field and 2 kilojoules of energy from a green laser.

"We were excited by the results," says ▶

► Mark Herrmann, director of the Z machine and the pulsed-power science centre at Sandia. “We look at it as confirmation that it is working like we think it should.”

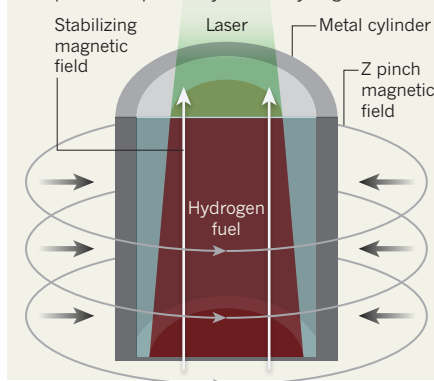
The experiment yielded about  $10^{10}$  high-energy neutrons, a measure of the number of fusion reactions achieved. This is a record for MagLIF, although it still falls well short of ignition. Nevertheless, the test demonstrates the appeal of such pulsed-power approaches to fusion. “A substantial gain is more likely to be achieved at an early date with pulsed power,” says nuclear physicist David Hammer of Cornell University in Ithaca, New York, who co-wrote a 2013 US National Research Council assessment of approaches to fusion energy.

With its relatively slim US\$5-million annual budget, MagLIF is a David next to two fusion Goliaths: the \$3.5-billion National Ignition Facility (NIF) at Lawrence Livermore National Laboratory in California, and the €15-billion (US\$20-billion) ITER experiment under construction in France. (Sandia has about \$80 million to operate the Z machine each year, but it serves other experiments in addition to MagLIF.) The NIF squashes fuel capsules using nearly 2 megajoules of laser energy, and ITER will use 10,000 tonnes of superconducting magnets in a doughnut-shaped ‘tokamak’ to hold a plasma in place to coax self-sustaining fusion.

Both of the big projects have run into problems. After a concerted two-year effort, NIF fell well short of achieving ignition by a 2012 deadline. Its fusion yields have since increased

### FEELING THE PINCH

Magnetized Liner Inertial Fusion uses a heating laser, a stabilizing magnetic field and a force called a Z pinch to implode a cylinder of hydrogen fuel.



markedly — nearly  $10^{16}$  neutrons were created in a recent shot, Herrmann says — but the more than \$300-million-a-year programme faces further budget cuts in 2014. Meanwhile, delays and budget overruns have become the norm at ITER. The facility is not expected to begin operations until 2027 — 11 years later than initially planned.

In addition to being cheaper, MagLIF seems to have technical advantages. The laser not only preheats the hydrogen fuel, but also makes it more conductive — and thereby more susceptible to the Z pinch. Furthermore, in a paper published late last year, MagLIF physicists showed evidence suggesting that the applied secondary

magnetic field, as well as insulating the fuel, may have the happy side effect of stabilizing the cylinder as it implodes (T. J. Awe *et al. Phys. Rev. Lett.* **111**, 235005; 2013). If so, that would cut down on hydrodynamic instabilities, which can disperse the energy and fuel before fusion can get going, says Stephen Slutz, a Sandia physicist who proposed the MagLIF system in 2009.

In the next few years, MagLIF scientists plan to turn up all three dials at their disposal. They can boost the Z machine to up to 27 million amperes; they can ramp up the magnetic field to as high as 30 tesla; and they plan to upgrade the laser to 8 kilojoules. They also aim to switch from fuel made of the hydrogen isotope deuterium to fuel containing both deuterium and another isotope, tritium — which should also lift yields. By 2015, they hope to achieve a yield of  $10^{16}$  neutrons, or about 100 kilojoules — enough to show that ignition is within reach.

It could be crucial to make progress quickly. The US National Nuclear Security Administration, the division of the Department of Energy that funds the NIF, the Z machine and other laser fusion efforts, plans to deliver an assessment to Congress in 2015 about the future of these technologies. If MagLIF hits its 100-kilojoule goal, it could bolster an argument for upgrading the Z machine to 60 million amperes or more, which simulations suggest would be sufficient to reach ignition.

“We’re all hoping that they will, in fact, find success with their early shots to justify the construction of a larger machine,” says Hammer. ■

SOURCE: SANDIA NATL LAB.

### CLIMATE CHANGE

# Water risk as world warms

First comprehensive global-impact project shows that water scarcity is a major worry.

BY QUIRIN SCHIERMEIER

When pondering the best way to study the impact of climate change, researcher Hans Joachim Schellnhuber liked to recall an old Hindu fable. Six men, all blind but thirsty for knowledge, examine an elephant. One fumbles the pachyderm’s sturdy side, while others grasp at its tusk, trunk, knee, ear or tail. In the end, all are completely misled as to the nature of the beast.

The analogy worked. Although many researchers had modelled various aspects of the global-warming elephant, there had been no comprehensive assessment of what warming will really mean for human societies and vital natural resources. But that changed last year when Schellnhuber, director of the Potsdam Institute for Climate Impact Research in Germany, and other leading climate-impact

researchers launched the Inter-Sectoral Impact Model Intercomparison Project. This aims to produce a set of harmonized global-impact reports based on the same set of climate data, which will for the first time allow models to be directly compared. Last month it published its initial results in four reports in *Proceedings of the National Academy of Sciences*<sup>1-4</sup>. These suggest that even modest climate change might drastically affect the living conditions of billions of people, whether through water scarcity, crop shortages or extremes of weather.

The group warns that water is the biggest worry. If the world warms by just 2 °C above the present level, which now seems all but

unavoidable by 2100, up to one-fifth of the global population could suffer severe shortages.

“Water and all that relies on it, from food to sanitation and public health, is an emblematic aspect of climate change whose urgency people tend to instantly understand,” says Schellnhuber.

To assess what a warmer world might mean for the human race, 30 groups from 12 countries have run thousands of simulations, using a standardized set of scenarios for greenhouse-gas emissions. They made projections of future water availability from a set of global hydrological models in conjunction with five state-of-the-art climate models<sup>1</sup> that combined projections of changes in temperature and precipitation with data on variables such as regional water cycles, river run-off and population.

The multi-model assessment suggests that, in vulnerable regions, climate change will significantly add to the problem of water scarcity

**“We are facing problems that result in domestic instability and migration.”**



that is already arising from population growth. The modellers found that climate-driven changes in evaporation, precipitation and runoff will result in a 40% increase in the number of people worldwide who must make do with less than 500 cubic metres of water per year — a commonly used threshold to signify ‘absolute’ water scarcity.

The spread between individual models was large — some suggested that global exposure to water scarcity will double; others predicted only modest change. But no matter what the spread, the greatest effects were seen between the present-day climate and a 2°C warmer world.

Despite the ambiguities, the exercise will make climate-risk analysis substantially more robust, says Johan Rockström, an expert on water resources at the University of Stockholm and director of the Stockholm Resilience Centre, who was not involved in the project.

“Impact models will never be able to provide the level of detail that ultimately matters for making a city or coastline climate-proof,” he says. “But they do serve as a first approximation to the severe problems deficient regions and nations are facing.”

Regions most at risk from water scarcity include parts of the southern United States, the Mediterranean and the Middle East. By contrast, India, tropical Africa and high latitudes in the Northern Hemisphere can expect to receive more water in a warming world.

The projected changes in water availability have knock-on effects in other areas that rely on water. The group that modelled the response of crops to climate change found negative impacts on yields of major crops in many agricultural regions<sup>2</sup>.

In addition, drought conditions are likely to become more frequent and severe in some parts



Water scarcity in parts of Africa could become worse, according to a complementary set of climate projections.

of South America, western and central Europe, central Africa and Australia, another project team reports<sup>3</sup>. Flood risk is less clear-cut, but river-flow simulations from global hydrology and land-surface models did show an increase in flood hazard in more than half of the world<sup>4</sup>.

Despite their uncertainty, the findings are “a stark reminder” that even moderate warming has the potential to cause severe natural, social and economic disruptions, says Rockström. “We are facing problems that result in domestic instability and migration.” Rethinking international trade with a view to giving the most needy nations better access to the global food market will be essential, he says.

Uncertainty, adds Schellnhuber, is no excuse for inaction. “Those who might say, ‘Come back when you’ve narrowed down the risk’ should be reminded that climate change is a treacherous gamble,” he says. “We don’t quite know the odds, but the chance of losing heavily might be a lot bigger than many tend to think.” ■

1. Schewe, J. et al. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1222460110> (2013).
2. Rosenzweig, C. et al. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1222463110> (2013).
3. Prudhomme, C. et al. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1222473110> (2013).
4. Dankers, R. et al. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1302078110> (2013).

## PHYSICS

# X-ray source left without home

*No plans to build next-generation accelerator despite large investment by US agency.*

BY EUGENIE SAMUEL REICH

Accelerator physicists have a vision: an energy-efficient X-ray source that can make high-resolution movies of molecules in chemical reactions. And the US National Science Foundation (NSF) has backed the dream — since 2005, it has invested more than US\$50 million to develop such a source, most likely beneath the campus of Cornell University in Ithaca, New York.

But there is one big problem: despite the inflow of cash, no US government agency has any plans to build the machine.

The source, called an energy recovery

linear accelerator (ERL), would be a hybrid of a synchrotron, in which electrons emit X-rays while whirling around a ring, and a free-electron laser, in which straight beams of electrons are induced to produce bright pulses of X-ray light.

The Cornell project is currently receiving \$27 million in a single award from the NSF’s materials division — by far the division’s largest grant for instrument development. But in July, the ERL concept was ranked the lowest of three potential next-generation X-ray sources by an advisory panel to the US Department of Energy. And in December, officials at the NSF told *Nature* that the agency has no plans

to move forward with construction.

Despite all this, Thomas Rieker, the NSF programme manager for the ERL materials grant, says that the research effort has been a success, providing component designs that would allow an accelerator to be built quickly. “We wanted to keep our options open,” he says. “That was the impetus for funding it.”

An NSF advisory panel had strongly recommended in 2008 that the NSF invest in an ERL. So why the turnaround? Agency officials now say that the NSF’s priorities and the budgetary climate have changed, and that a machine costing upwards of \$1 billion would not be a good use of taxpayers’ money. ►

**World's largest  
selection of  
validated ELISAs**



**Guaranteed  
quality at  
unbeatable prices**

- ISO 13485 certified
- Rigorously tested
- 100% guaranteed
- Thousands of citations
- Pre-coated strip plates
- Manufactured in-house

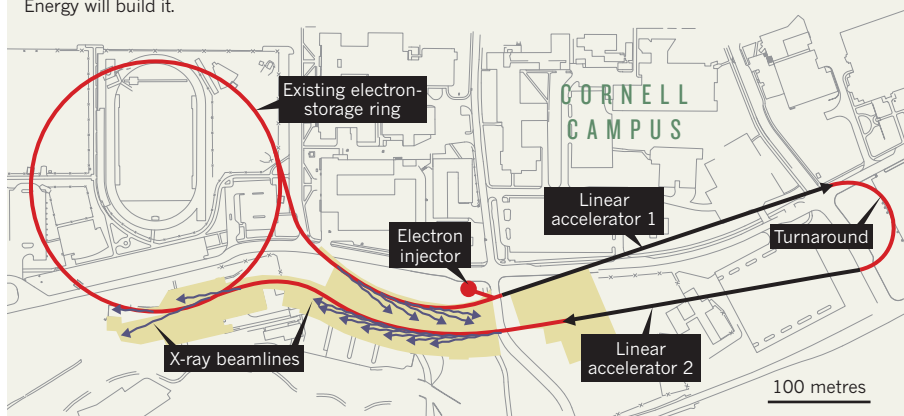
Browse  
RayBio®  
ELISAs



[www.raybiotech.com](http://www.raybiotech.com)

## ELECTRONS IN RECOVERY

Cornell University in Ithaca, New York, wants to convert an existing synchrotron into an energy recovery linear accelerator, but it is unlikely that either the US National Science Foundation or the Department of Energy will build it.



SOURCE: CORNELL UNIV.

► Some physicists are expressing frustration over seeing so much research money apparently going nowhere. “The NSF should really decide if there’s a real need for this in the country,” says Sunil Sinha, a condensed-matter physicist at the University of California, San Diego, who advised on the energy-department panel.

The idea for an ERL was developed in 1965 by Cornell physicist Maury Tigner. It involves injecting electrons into a linear accelerator (linac) and then wiggling the particles to prompt the emission of X-ray pulses. The energy-recovery aspect comes from a loop that ushers the electrons gently around to enter the linac a second time. Their arrival is timed so that their energy is transferred to a new bunch of electrons that will then be accelerated.

The approach has several advantages. For starters, it would be vastly more energy-efficient than a free-electron laser, which recovers no energy. That makes it practical to keep electrons streaming continuously, rather than in widely separated bunches. An ERL can also focus its electron beam, and hence the resulting X-rays, to a tighter spot than the beams in current synchrotron rings, which spread out as they lose energy going around in circles. This would allow for more-advanced studies of the atomic energy levels in materials.

Japan and the United Kingdom have both expressed interest in building an ERL, and there is a small demonstration version of an infrared ERL at the Thomas Jefferson National Accelerator Facility in Newport News, Virginia. But Cornell’s plan is the most advanced X-ray ERL effort in the United States.

Grant documents stress that the ERL research is not site-specific, meaning that it could feed into projects elsewhere. But most experts think that, if an ERL gets built, it would be at Cornell, where it could reuse the tunnels

of an existing NSF-funded X-ray light source, the Cornell High Energy Synchrotron Source (see ‘Electrons in recovery’). “We feel the construction of an ERL can go right ahead,” says Georg Hoffstaetter, an accelerator physicist who is leading the Cornell effort.

The capabilities of an ERL would overlap with those of other planned light sources. In California, the Department of Energy has plans to build a free-electron laser, perhaps by upgrading the Linac Coherent Light Source at the SLAC National Accelerator Laboratory in Menlo Park, California. This machine would provide images of materials with unprecedented resolution in space and time, using fast pulses of high-energy X-ray beams.

Pulses of X-ray light from an ERL would not be as fast, but they would be gentler, and nearly continuous — more appropriate for probing sensitive samples such as biological specimens. However, next-generation ring-shaped light sources, such as an upgrade planned at the Advanced Photon Source near Chicago, Illinois, will also stream continuous light. Although less bright and lower in energy than an ERL, such sources would still prove useful for biological imaging.

The energy department’s decision to go with the other machines will make it harder for the ERL to justify itself scientifically, says Paul Evans, a materials scientist at the University of Wisconsin–Madison. “Defining the niche they’re headed for is the critical challenge.”

Even if the ERL is not built, Cornell scientists say that the research has been useful. Their work is aiding design of superconducting cavities for the Department of Energy’s future free-electron laser, which could also one day have energy-recovery loops tacked on. Cornell has also developed a high-current electron gun that could be used in other accelerators to generate X-rays or to study particle collisions.

But although he takes satisfaction in the spin-off possibilities, Hoffstaetter is not ready to give up on the ERL’s construction. “The ERL is a wise investment,” he says. ■

**“The NSF should really decide if there’s a real need for this.”**



## RESEARCH

# What to expect in 2014

*Nature takes a look at what is in store for science in the new year.*

## TRANSGENIC MONKEYS

Several research groups, including a team led by geneticist Erika Sasaki and stem-cell biologist Hideyuki Okano at Keio University in Tokyo, hope to create transgenic primates with immune-system deficiencies or brain disorders. This could raise ethical concerns, but might bring us closer to therapies that are relevant to humans (mice can be poor models for such disorders). The work will probably make use of a gene-editing method called CRISPR, which saw rapid take-up last year.

## SPACE PROBES

The European Space Agency's Rosetta spacecraft could become the first mission to land a probe on a comet. If all goes well, it will land on comet Churyumov-Gerasimenko in November. Mars will also be a busy place: India's orbiter mission should arrive at the planet in September, about the same time as NASA's MAVEN probe. And NASA's Curiosity rover should finally make it to its mission goal, the slopes of the 5.5-kilometre-high Aeolis Mons, where it will look for evidence of water. Back on Earth, NASA hopes to launch an orbiter to monitor atmospheric carbon dioxide.

## NEURAL FEATS

Neurobiologist Miguel Nicolelis at Duke University in Durham, North Carolina, has developed a brain-controlled exoskeleton that he expects will enable a person with a spinal-cord injury to kick the first ball at the 2014 football World Cup in Brazil. Meanwhile, attempts are being made in people with paralysis to reconnect their brains directly to paralysed areas, rather than to robotic arms or exoskeletons. In basic research, neuroscientists are excited about money from big US and European brain initiatives, such as Europe's Human Brain Project.

## NOVEL DRUGS

In the pharmaceutical industry, all eyes are on trial results from two competing antibody treatments that harness patients' immune systems to fight cancer. The drugs, nivolumab and lambrolizumab, work by blocking proteins that prevent a person's T cells from attacking tumours. In early tests, the drugs evoked a better level of response in patients

than ipilimumab, a similar therapy that was launched in 2011 to treat advanced melanoma.

## RENEWABLE REVOLUTION

Semiconductors known as perovskites convert light energy into electricity. They are cheap to build and have already shown conversion rates of more than 15% (a leap from 4% when the feat was first reported in 2009). Expect to see still-higher efficiencies this year, perhaps

to release the first data from a disposable sequencer the size of a memory stick, which it is sending to scientists for testing. It promises to read longer strands of DNA than other techniques (potentially useful in sequencing mixed samples of bacterial DNA, for example), and to show results in real time.

## A BETTER CLIMATE

The Intergovernmental Panel on Climate Change will complete its fifth assessment report by November. The findings of working groups II and III will focus on the impacts of climate change, and on how societies can adapt to or mitigate those effects (working group I published its findings last year). Away from formal negotiations, United Nations secretary-general Ban Ki-moon is hoping for "bold pledges" on emissions at a summit in New York in September. In research, a large carbon capture and storage project in Canada — the Can\$1.24-billion (US\$1.17-billion) Boundary Dam coal power-plant in Saskatchewan — begins commercial operation in April.

## MAKING WAVES

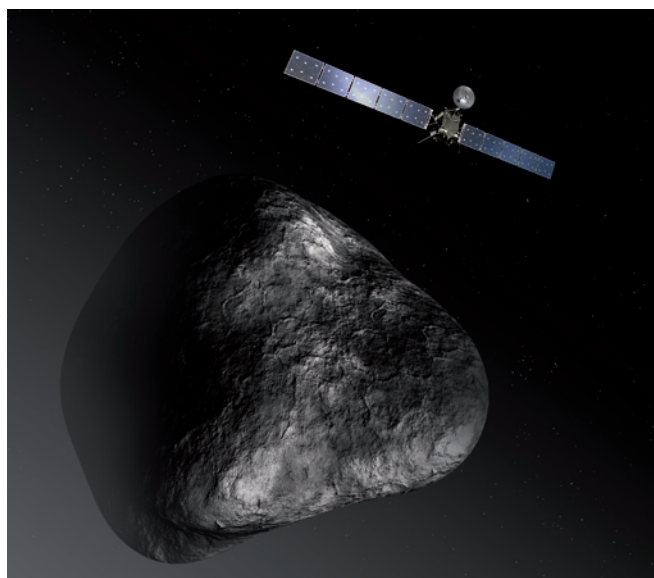
The European Space Agency's Planck satellite team should release data on how the polarization of photons from the Universe's

cosmic microwave background varies across the sky. This esoteric pattern is thought to have been generated by 'inflation', the rapid expansion of the Universe after the Big Bang. If it can be detected, its details could provide evidence of relic gravitational waves, thought to have perturbed space-time in the early Universe.

## STEM-CELL REGENERATION

A Japanese team will start the first clinical trials using induced pluripotent stem cells this year — but don't expect results anytime soon. And biotechnology firm Advanced Cell Technology in Santa Monica, California, says that it will release data from two trials using human embryonic stem cells — the only two to gain approval from US drug regulators. These two studies involve injecting stem-cell-derived retinal cells into the eyes of around 30 people with one of two forms of non-treatable degenerative blindness. ■

COMPILED BY RICHARD VAN NOORDEN



An artist's impression of the European Space Agency's Rosetta probe, which aims to be the first to land on a comet.

reaching 20% — the same as the lower end of existing commercial silicon-based photovoltaics. A team at the University of Oxford, UK, also hopes to make lead-free perovskites.

## HIV HOPE

In 2013, two research teams showed that 'broadly neutralizing' antibodies that target an array of HIV types quickly cleared an HIV-related virus in monkeys. The therapy will be tested in people who carry HIV, with results expected in the autumn. Meanwhile, last year's curing of a baby born with the virus might lead to wider trials of the technique used: high doses of antiretroviral drugs given at birth.

## MINIATURE SEQUENCER

Technology that rapidly sequences DNA as it is fed through a ring of proteins, known as a biological nanopore, will hit the market this year after decades of development. Oxford Nanopore Technologies in Oxford, UK, aims

# THE ACCIDENTAL EPIGENETICIST

BY STUDYING DISADVANTAGED CHILDREN, RICHARD TREMBLAY HAS TRACED THE ROOTS OF CHRONIC AGGRESSIVE BEHAVIOUR BACK AS FAR AS INFANCY. NOW HE HOPES TO GO BACK FURTHER.

BY STEPHEN S. HALL

**H**ochelaga was the original Iroquoian name for the village that ultimately became Montreal, but it is also the name of a rough-hewn French-Canadian neighbourhood located east of — and a world away from — the cosmopolitan city centre. The district's tidy two- and three-storey brick duplexes, adorned with Montreal's characteristic wrought-iron staircases, predominantly house families that have, because of poverty and lack of education, never quite attained thriving middle-class status.

During the 1980s, public-school officials identified Hochelaga and many other impoverished neighbourhoods in the eastern part of Montreal as places where kindergarten children disproportionately displayed severe behavioural problems, such as physical aggression. The school system asked a young University of Montreal psychologist named Richard Tremblay for help.

"Their parents didn't have a high-school diploma, and many of the mothers had their first child before the age of 20," Tremblay says of the families he began to study, as he walks along Rue Ontario in Hochelaga on a sunny afternoon in September. Those were

the women, he adds, "most at risk of having children who have problems".

Over the past three decades, Hochelaga and similar neighbourhoods have served as living laboratories in the study of the roots of aggression. Since 1984, Tremblay and his collaborators have followed more than 1,000 children from 53 schools in the city from childhood into adulthood. And in 1985, he initiated a groundbreaking experiment in which some families of at-risk children were given support and counselling to help curb bad behaviour. His research overturned ideas about when aggressive behaviour first emerges, and showed that early intervention can deflect children away from adult criminality.

The idea that a nurturing environment provides better outcomes for children hardly qualifies as news, but Tremblay has taken this idea in a provocative direction in the past ten years. He has joined researchers at McGill University in Montreal and the US National Institutes of Health (NIH) in Bethesda, Maryland, to investigate how nurturing or adverse environments might exert their effects at the molecular level, influencing gene expression through a mechanism known as epigenetics. Tremblay's Canadian cohorts are part







ROGER LEMOYNE/REDUX PICTURES/EVINE

of a growing trend for using longitudinal studies, which follow the same individuals over an extended period of time, to look for epigenetic signatures that might affect health and behaviour later in life. Research in this area is still preliminary — and not without its critics — but Tremblay believes that a firm grasp of early epigenetic effects could inform interventions to influence everything from obesity to mental illness.

“There is a body of evidence, from natural experiments and actual experiments, showing that early-life experiences affect long-term outcomes such as crime, health and wages,” says James Heckman, a Nobel-prizewinning economist at the University of Chicago in Illinois who is currently working with Tremblay on an early-intervention study with at-risk pregnant mothers in Dublin. The work of Tremblay and others, he says, “has established a firmer biological basis for how early-life experiences affect these processes”.

### SETTING GOALS

Tremblay’s own early life revolved around sport. His father, Wilfrid Tremblay, played Canadian football from 1938 to 1951, and Richard was an accomplished ice-hockey goalie. When Jacques Plante, the Hall of Fame goalie for the Montreal Canadiens, suffered an injury during the Stanley Cup Playoffs in 1961, a team representative called the then-17-year-old Tremblay asking if he could report to the minor league practice rink the next morning. Tremblay, soft-spoken and mild-mannered, allows that he was “invited to join” the most illustrious franchise in Canadian sports, but concluded that he was too small to play at the professional level. He decided to attend college instead.

Tremblay studied physical education at the University of Ottawa. But before his final year, he read a cult novel by J. R. Salamanca called *Lilith* (Simon & Schuster, 1961), about a recreational therapist who falls in love with a young female patient at a psychiatric hospital. To a naive 20-year old, the work sounded fascinating, and when he returned to college that autumn he applied for a job as a recreational therapist at a high-security psychiatric hospital in Joliette, Quebec. He quickly found himself in over his head, working with convicted murderers and violent criminals. But it was during this time that he first started to wonder about the psychology of aggression. “It shows how a novel can change a life,” he says.

The hospital agreed to send him to get a master’s degree in psychology, which he pursued at the University of Montreal. As Tremblay likes to say: “The first thing I did after finishing my master’s degree was to go to jail for three years.” That was the Pinel Institute, a new maximum-security psychiatric hospital in Montreal. Most of the people there, he says, “had killed someone or were dangerous to the point of killing themselves, or others”. Despite



the danger, he found himself going to work on his days off to play sports with the residents. “I loved it,” he recalls.

Then, in 1971, the University of Montreal decided to create a school focused on children with behavioural problems. The university wanted to hire Tremblay, one of the most promising students to come out of its psychology programme, to join the faculty. But he needed a PhD first, so the university paid for his training at the University of London’s Institute of Education.

That turned out to be a defining experience. Tremblay arrived in London with a sheaf of Rorschach blots and a grounding in psychoanalysis, but there he was exposed to the ‘longitudinal’ philosophy of pioneering human-growth biologist James Tanner, child psychiatrist Michael Rutter and others. He came away with a lesson that has informed the rest of his scientific career: the best way to study any aspect of human development is to conduct longitudinal studies. He threw away his Rorschach blots and, in the late 1970s, headed back to Montreal.

### AGGRESSIVE START

By then, Tremblay was eager to launch his own longitudinal study. He got his chance in the early 1980s. School officials came to him with the problem of hyperactive, physically aggressive kindergarten boys. He had never worked with children before and never imagined doing so, but he recognized it as an opportunity to explore the origins of aggressive behaviour. “The idea became very clear,” he says. A longitudinal study of kindergarten children would give him a chance to link childhood behaviours with adolescent and adult outcomes.

In 1984, he started tracking boys from dozens of schools. Funding was initially provided for three years, but nearly three decades later Tremblay and his colleagues continue to follow many of the men involved. They have published more than 160 papers on the group.

Just one year in, when the boys were seven years old, Tremblay obtained a grant to add a randomized, controlled experimental intervention. Teams of four psychologists would visit the families of about 50 boys every two weeks. They counselled parents on identifying and correcting aggressive behaviour, and trained teachers to do the same. In addition, they attempted to socialize unruly boys, and they integrated problematic boys with well-behaved children to provide positive peer role models.

The Montreal intervention began at a time known informally among criminologists as the ‘nothing works’ era, when there was widespread pessimism about the potential to rehabilitate juvenile delinquents and adult criminals. Tremblay’s intervention was

labour-intensive and extremely expensive, and he recalls fretting that he was spending millions of dollars on a study but might end up with nothing to show for it. “I guess I lost hope — in working with juvenile delinquents — that we could make a difference,” he says.

The intervention lasted about two years, but the results would take much longer to become apparent. One of the first people to see hints that it was working was Joan McCord, a criminologist at Temple University in Philadelphia, Pennsylvania. McCord had a reputation for ferreting out data that challenged conventional wisdom, most notably when she demonstrated

behaviour in early childhood that continued into their teens. The roots of physical aggression — and, by extrapolation, the origins of violent behaviour later in life — lie before the age of six, says Nagin. That is, before Tremblay’s kindergarten cohort even began.

Even as Nagin and Tremblay were analysing the original Montreal data, Tremblay had begun another longitudinal study designed to look at aggression before kindergarten. It was a birth cohort based in Quebec, and the resulting data suggested that aggressive behaviour was evident at 17 months and peaked at around 42 months<sup>4</sup>. This and later work culminated in Tremblay’s ‘original sin’ hypothesis: that physical aggression is the default setting in human behaviour<sup>5</sup>. It peaks between the ages of two and four, and is usually socialized out of children by the time they enter school (see ‘Aggression regression’). “We took the view that violence, and physical aggression, is a part of us as a species,” says Nagin, “so the issue is not how we learn it, but rather how we learn to control it.”

Many criminologists dismissed the findings. They argued not that the idea was wrong, but that it was irrelevant — that chronic childhood aggression is trivial compared with murder and rape in adulthood, and that the former does not explain the latter. Most still focus primarily on delinquency during adolescence, and for good reason, says sociologist Robert Sampson at Harvard University in Cambridge, Massachusetts. “Early childhood is centrally important, but it’s not determinative, because there are still changes [in behaviour] later on.”

Yet the Montreal and similar longitudinal studies show that heightened physical aggression at a young age correlates with serious antisocial behaviour in adolescence and adulthood, says Tremblay. He is fond of citing the view that Saint Augustine offered some 1,600 years ago: “It is not the infant’s will that is harmless,” he wrote, “but the weakness of infant limbs.”

Yet the Montreal and similar longitudinal studies show that heightened physical aggression at a young age correlates with serious antisocial behaviour in adolescence and adulthood, says Tremblay. He is fond of citing the view that Saint Augustine offered some 1,600 years ago: “It is not the infant’s will that is harmless,” he wrote, “but the weakness of infant limbs.”

### MARKING TIME

With Saint Augustine’s headstrong infants in mind, Tremblay increasingly pondered the effects of the environment at earlier and earlier ages. Like many researchers studying behaviour, he had looked into what role genes might have in aggression, but he was dissatisfied. Genetics did not tell the whole story. Tremblay was primed, therefore, to hear about the work of Moshe Szyf, a cancer biologist at McGill, at a small Vancouver meeting in 2004.

Szyf had been tracking the addition and removal of methyl groups to DNA, which can silence or activate genes. Scientists were interested in whether these methylation marks

THE ISSUE IS NOT HOW  
WE LEARN VIOLENCE,  
BUT RATHER HOW WE  
LEARN TO CONTROL IT.

in the 1970s that a famous US longitudinal experiment — the Cambridge–Somerville Youth Study, in which juvenile delinquents were mentored and supported — had actually harmed the young men it had aimed to help<sup>1</sup>. Conversely, the Montreal intervention seemed to be working as intended. With each follow-up assessment, boys in the intervention arm displayed not only less delinquent behaviour than controls, but also better school performance, lower consumption of drugs and alcohol, and better social skills.

Data gathered 15 years after the intervention ended revealed that it produced persistent positive effects. The boys whose families received support had a 46% graduation rate as opposed to 32% for controls. And, at the age of 24, fewer of them had criminal records: 22%, versus 33% for controls<sup>2</sup>.

But Tremblay wasn’t just seeking ways to mitigate bad behaviour — he was looking to uncover where it began. In the mid-1990s, he began to collaborate with Daniel Nagin, a criminologist at Carnegie Mellon University in Pittsburgh, Pennsylvania. Nagin applied a more sophisticated statistical metric to the burgeoning Montreal data set. The results, published in 1999, made it clear that the trajectory towards antisocial behaviour and criminality in adolescence begins very early in life<sup>3</sup>. Most children exhibit decreasing aggression between the ages of 6 and 15: they learn to control their aggressive impulses. Only about 4% of the boys displayed highly aggressive

might allow the environment to influence gene expression over an organism's lifetime. Michael Meaney, a developmental neurobiologist also at McGill, collaborated with Szyf to show that newborn rat pups generously licked and groomed by their mothers had different patterns of DNA methylation from those that received less maternal attention<sup>6</sup>. These changes reached the brain, where the methylation pattern altered the activity of a gene that plays a central part in the animal's response to environmental stress. Maternal nurture, Szyf argued, was a form of 'environmental programming' that altered the activity and function of genes in ways that persisted throughout life.

For Tremblay, it was "as if the roof blew off" the room. The McGill experiments suggested a biological explanation for what he had been tracking for 20 years. As he walked to dinner with Szyf that evening, Tremblay pressed for a possible collaboration.

Human studies of this sort were uncharted territory. So Tremblay initiated a parallel line of animal research with Stephen Suomi, who heads the primate laboratory at the NIH's Eunice Kennedy Shriver National Institute of Child Health and Human Development in Bethesda. Both scientists had noted behavioural similarities between the chronically aggressive, hyperactive boys in the Montreal study and a group of aggressive monkeys that Suomi had raised under conditions of early maternal deprivation. Tremblay, Suomi and Szyf set out to run DNA-methylation studies on two sets of monkeys: a group nurtured by their mothers, and another deprived of maternal nurturing from shortly after birth. It took nearly a decade of difficult molecular-biology work headed up by Nadine Provençal at McGill, but in the past year or so, the researchers have begun to publish their findings.

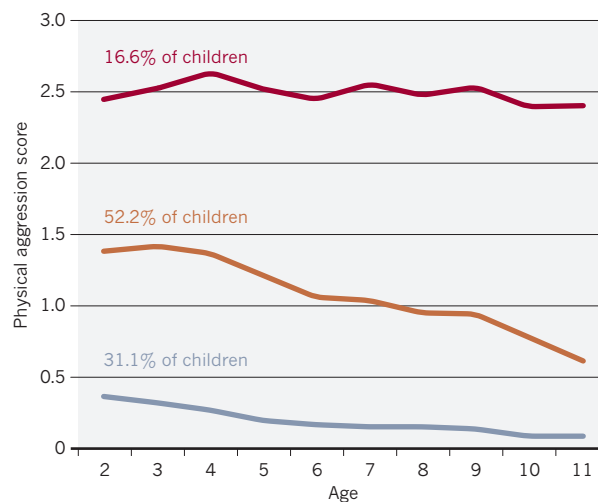
The first primate study found distinct differences in DNA-methylation patterns between nurtured monkeys and those separated from their mothers<sup>7</sup>. The epigenetic residue of post-natal adversity was broad, according to Suomi, affecting more than 4,000 genes — about one-fifth of the genome — and tending to cluster in certain chromosomal regions. Moreover, the epigenetic modifications seemed to alter expression of a gene that Suomi's group had shown to be crucial to the function of the neurotransmitter serotonin<sup>8</sup>, low levels of which have been associated with elevated stress and aggression in humans. "These are not random changes," Suomi says. "They follow particular pathways." The marks remained stable in monkeys up to 8 years old — an age roughly equivalent to 30 in humans.

Although the team was able to test both

brain and white blood cells from the monkeys, they only had access to blood from the men of the Montreal cohort. Even so, studies are starting to offer a complementary human picture. In July, Szyf and Tremblay reported that men with a history of chronic aggression dating back to kindergarten had significantly lower blood levels of immune molecules called cytokines than normal controls from the cohort<sup>9</sup>. These molecules are typically

## AGGRESSION REGRESSION

A study of more than 10,000 Canadian children pointed to three basic trajectories for physical aggression. Most become less aggressive between the ages of 2 and 11 years, but a minority maintain a high level of aggression throughout childhood.



activated during the body's response to stress, and animal studies have demonstrated a link between aggression and lower levels of a cytokine called interleukin-6, which was also lower in the chronically aggressive men. In a second study, Szyf and Tremblay showed that members of the Montreal Longitudinal Study with a long-standing history of aggression had a distinctly different pattern of DNA methylation in the genes encoding the cytokines, compared to men with a less aggressive behavioural profile<sup>10</sup>.

The early human research has its shortcomings. For starters, the sample size is very small: only seven males with a history of aggression could be tracked down from the cohort for testing, along with 25 controls. And white blood cells are by no means the same as neurons, although Suomi notes that there is considerable overlap between the methylation patterns of the two cell types in the primate studies. Moreover, many researchers remain cautious about recent human epigenetic studies. Attributing behavioural consequences to DNA methylation may be overreaching, says Adrian Bird, a geneticist at the University of Edinburgh, UK. "These are all correlations," he says, "and often the magnitude of the change is very small indeed."

Tremblay is the first to admit that the story

is far from simple: hundreds of genes are involved, and any single expression change is probably subtle. Yet, he says, "it seems relatively clear that there are large differences in DNA methylation between those who have a history of chronic aggression compared to those who have normal development". He is convinced that the benefits of nurture merit early intervention programmes, regardless of the uncertainties in the biological part of

the story. And he thinks that earlier intervention may produce even better results. "If we support these parents during pregnancy and if we help these women have a better lifestyle during pregnancy, with less stress, it should affect brain development, and these children should be better able to learn how to control their aggressive behaviour," he says.

He is already testing that hypothesis. In 2007, he accepted a ten-year appointment at University College Dublin, where he is assisting on several early-childhood longitudinal studies. One, called Preparing for Life and headed by economist Orla Doyle, is testing a preventive intervention in 200 pregnant women from a disadvantaged area of north Dublin. During their pregnancies, the women received intensive home visits covering everything from nutrition, smoking, alcohol and drug counselling to support in marital relationships. And the support continues until the children reach the age of four. James Heckman, who is also collaborating on the study, says that the plan includes future epigenetic studies of the cohort.

"To solve the aggression problems, which are mainly a male problem, we need to focus on females," Tremblay says. "If you ameliorate the quality of life of women, it will transfer to the next generation." ■

**Stephen S. Hall** is a science writer in New York and teaches public communication to graduate students in science at New York University.

1. McCord, J. *Am. Psychol.* **33**, 284–289 (1978).
2. Boisjoli, R., Vitaro, F., Lacourse, E., Barker, E. D. & Tremblay, R. E. *Brit. J. Psychiatry* **191**, 415–419 (2007).
3. Nagin, D. & Tremblay, R. E. *Child Dev.* **70**, 1181–1196 (1999).
4. Côté, S. M. et al. *Arch. Gen. Psychiatry* **64**, 1305–1312 (2007).
5. Tremblay, R. E. *J. Child Psychol. Psychiatry* **51**, 341–367 (2010).
6. Weaver, I. C. G. et al. *Nature Neurosci.* **7**, 847–854 (2004).
7. Provençal, N. et al. *J. Neurosci.* **32**, 15626–15642 (2012).
8. Bennett, A. J. et al. *Mol. Psychiatry* **7**, 118–122 (2002).
9. Provençal, N., Suderman, M. J., Vitaro, F., Szyf, M. & Tremblay, R. E. *PLoS ONE* **8**, e69481 (2013).
10. Provençal, N. et al. *PLoS ONE* **8**, e71691 (2013).



# COMMENT



**CULTURE** This year's hot tickets, from opera to theatre, museums to movies **p.22**

**PHYSICS** A compelling personal manifesto for the multiverse, on four levels **p.24**

**EDUCATION** Massive open online course for dementia carers reaches new learners **p.26**

**OBITUARY** Frederick Sanger, winner of two Nobel prizes, remembered **p.27**

CHEN-MIN CHUNG/IN PICTURES/CORBIS



A worker sorts through stripped computer boards in Guiyu, China.

## Track flows to manage technology-metal supply

Recycling cannot meet the demand for rare metals used in digital and green technologies, says **Andrew Bloodworth**. A more holistic approach is needed.

**D**emand for metals is soaring as the global population booms and millions of people in emerging economies aspire to a Western lifestyle. The variety of metals we use has also expanded as technology has advanced. As a result, historic fears regarding metal scarcity and resource depletion have returned in the past ten years.

Concerns focus on the future supply of metals such as indium, lithium, rare-earth elements, tellurium and germanium, all of which are crucial to delivering new digital and low-carbon energy technologies, including photovoltaics and electric cars.

In 2009, the issue came to global prominence when China reduced its exports of

rare-earth elements, as the government sought to maintain supply to its rapidly expanding domestic manufacturing sector.

Geopolitical and socio-economic risks — such as territorial disputes in Asia or labour relations in southern Africa — can interrupt supply because technology metals are produced in very few locations. Commercial barriers compound the issue. Investment in these materials can be risky because they are difficult to extract and the markets are small, complex and volatile compared with those for iron, copper and aluminium.

To secure supplies of metals for future technology, the scientific, industrial and policy communities must work together. The

numerous assessments that governments have commissioned fall short. They identify key issues, but generate lots of sterile argument as to whether a particular metal is 'critical'. The solutions they point to are generic and of little practical use.

Prominent among these broad-brush responses has been the implication that the security of technology-metal supply in mainland Europe and the United Kingdom can be achieved mostly through recycling. Although recycling is important for managing stocks of common industrial metals, its application to technology metals is more complex. Some materials are impractical or impossible to retrieve after use. ►



► More primary sources will be needed to meet rising demand and replace lost technology metals. To find new resources, the geological processes that concentrate these metals need to be better understood. And to increase efficiency and avoid unintended environmental impacts, the flows of individual metals need to be mapped from the ground to the end of their use.

### RARE RESOURCES

Demand for technology metals has exploded in the past 40 years, with 80% of the cumulative global production of gallium, rare-earth elements, platinum-group metals and indium taking place since 1980 (ref. 1). Growth is expected to continue for the foreseeable future<sup>2</sup>.

Most technology metals are mined in only a few places. In 2011, for example, 72% of global cobalt came from the Democratic Republic of the Congo<sup>3</sup> and 57% of indium originated from China (see [go.nature.com/crtooz](http://go.nature.com/crtooz) and 'Metal producers'). Such metals are produced in low quantities. In 2011, just 72,900 tonnes of tungsten was extracted globally, compared with 45.2 million tonnes of aluminium and 1.5 billion tonnes of crude steel<sup>3</sup>.

Some studies have concluded that scarcity and depletion of technology metals are unavoidable as rising consumption will exceed current reserves<sup>4</sup>. These apocalyptic forecasts fail to take into account that geological reserves are dynamic, expanding as metal prices rise and extraction of lower-grade ores becomes economical and tractable, and contracting as prices fall. A combination of price pressures and technical advances has kept global reserves of most metals steady or growing over the past 50 years<sup>5</sup>.

Because technology metals were of limited economic interest until recently, there has been little imperative to look for them. Consequently, not much is known about their distribution in Earth or the natural processes that concentrate them.

As knowledge improves, we will be able to reappraise old mining areas and explore new frontiers. Former mines in southwest England might hold promise for tungsten, for example, and a significant deposit of heavy rare-earth elements was identified in 2009 at Norra Kärr in Sweden. However, overcoming public objections to new mines can be a major challenge, especially in the developed world, where populations are often reluctant to accept the resource consequences of conspicuous consumption.

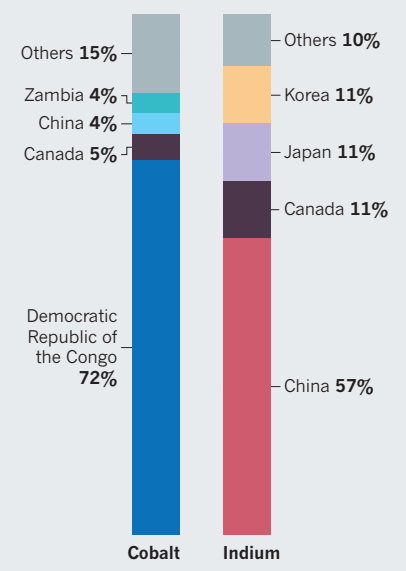
### RECYCLING IS NOT ENOUGH

Secondary metals, recycled from defunct products, provide valuable supplementary resources. But secondary stock will never meet growing demand. And recycling has technical limits.

From mobile phones to motor vehicles,

### METAL PRODUCERS

In 2011, most cobalt came from the Democratic Republic of the Congo, and most indium from China. Mined amounts have risen rapidly in the past 20 years owing to demand from technologies.



technology metals are used in myriad applications. Up to 60 different elements go into the manufacture of microprocessors and circuit boards<sup>6</sup>, usually in tiny quantities and often in combinations that are not found in nature.

Whether a metal can be recovered once a device is defunct depends on the element's value, concentration and accessibility when it is combined with other materials<sup>7</sup>. Precious metals — platinum-group metals and gold — are the main target in the processing of used circuit boards. Lower-value copper, antimony and indium can be recovered at the same time. But metals such as tantalum, gallium, germanium and rare-earth elements are oxidized and effectively lost in the smelter slag<sup>1</sup>.

Recycling technology metals is most economically attractive when they are highly concentrated, for example in manufacturing scrap. Around 70% of the indium used in producing flat-screen displays, for example, finds its way into scrap, which is then recycled<sup>8</sup>.

To fix bottlenecks and inefficiencies requires measuring technology-metal stocks and understanding how they flow through the whole supply chain — from mining to concentration, extractive and process metallurgy, manufacturing, use, reuse, recycling, dispersal and disposal<sup>9</sup>. For instance, improving recovery technology at tungsten mines would increase the amount of the metal in the ore that ends up reaching the smelter (just 75% for tungsten, in contrast to 90% for gold).

In theory, more than 90% of platinum-group metals used in autocatalysts can be recovered. In practice, only 50–60% is retrieved from European scrap cars because many vehicles are exported second-hand to

places that lack recycling facilities. Analysis of metal flows could show whether a scheme to retrieve lost catalytic converters would be more effective than another type of scheme, such as a 2011 proposal by a UK waste-management company to recover these metals from road sweepings. Autocatalysts contain about 0.2% platinum-group metals; sweepings contain less than 1 part per million<sup>10</sup>.

Addressing technological barriers to resource efficiency in this way is a focus of initiatives, such as the European Innovation Partnership on Raw Materials — a network of European countries aimed at increasing the availability of raw materials across the region.

However, mapping the life cycle of critical metals is challenging. The volumes are low; extraction, processing and recycling are handled by just a few organizations; and commercial confidentiality can make data and contacts hard to find.

### ONE SYSTEM

In the past five years, concerns over securing supplies of technology metals have evolved from near-panic over physical depletion and Chinese geopolitical muscle-flexing, to a dangerous assumption by some policy-makers that recycling is the panacea. A more holistic approach is needed.

Primary and secondary sources must be considered as part of one system that needs to be wholly understood. Basic statistical data are crucial. Better dialogue between producers, processors, consumers and recyclers will be needed. Policy-makers must assess how technology metals are used and combined, and the impact this has on the economic and environmental viability of recycling them.

The benefits of securing supplies of technology metals are clear. Improving the efficiency and reducing the environmental footprint of extraction and processing of these metals from primary sources is a major opportunity for industry and researchers. ■

**Andrew Bloodworth** is science director for minerals and waste at the British Geological Survey, Nottingham, UK.  
e-mail: [ajbl@bgs.ac.uk](mailto:ajbl@bgs.ac.uk)

- Hagelüken, C., Drielsmann, R. & Ven den Broeck, K. in *Precious Materials Handbook* (eds Sehr, U. & Grehl, M.) Ch. 1 (Hanua-Wolfgang, 2012).
- Graedel, T. E. & Erdmann, L. *Mat. Res. Soc. Bull.* **87**, 325–331 (2012).
- British Geological Survey *World Mineral Production 2007–2011* (BGS, 2013).
- Ragnarsdóttir, K. V. *Nature Geosci.* **1**, 720–721 (2008).
- Crowson, P. C. F. *Mineral Econ.* **24**, 1–6 (2011).
- Gunn, A. G. in *Proc. 12th Bienn. Soc. Geol. Appl. Miner. Depos. Meet.* (SGA, 2013).
- Reck, B. K. & Graedel, T. E. *Science* **337**, 690–695 (2012).
- Jackson, B. & Mikolajczak, C. *Availability of Indium and Gallium* (Indium Corporation, 2012).
- Allwood, J. M. & Cullen, J. M. *Sustainable Materials* (UIT Cambridge, 2012).
- Jackson, M. T., Sampson, J. & Pritchard, H. M. *Sci. Total Environ.* **385**, 1–3 (2007).



Inspired by a solar eclipse: *Prince Igor*, coming to the New York Metropolitan Opera in February.

## CULTURE

# Top tickets for 2014

It promises to be a heady year for science in culture: fans can steep in the sumptuous world of colour, unpeel the upside of failure, explore neural pathways, revisit the First World War, mend a rip in space-time, go pterosaur-spotting and traverse a mammoth-ridden nation. **Daniel Cressey** investigates.

## Prince Igor

*Metropolitan Opera House, New York*  
Premiere 6 February

The cosmic quest to comprehend the Universe has provided rich metaphorical pickings for drama, as Bertolt Brecht's play *Life of Galileo* and Philip Glass's opera *Kepler* attest. Nineteenth-century Russian composer, chemist and physician Alexander Borodin also looked to the skies — specifically, a solar eclipse — for inspiration. The result is the opera *Prince Igor*, based on the epic story of the twelfth-century monarch. Playing at New York's Metropolitan Opera House for the first time in almost a century, the production is directed by Dmitri Tcherniakov and conducted by Gianandrea Noseda. Ildar Abdrazakov stars in the title role.

## Fail

*Dublin Science Gallery*  
7 February – 27 April

Should we do more to celebrate failure? Contributors to this exhibition at the Dublin Science Gallery would like us to. Curated by Jane Ni Dhulchaointigh, the Irish inventor of the silicone rubber Sugru — used for everything from fixing to sculpting — the show features 21 objects selected by household names in fields including science and engineering. The items personify the failure survived or exploited by these individuals on their way to success. Be prepared to re-examine flops and losing streaks and, as Irish playwright Samuel Beckett had it, to learn how to “fail better”.

## Britain: One Million Years of the Human Story

*Natural History Museum, London*  
13 February – 28 September

About 900,000 years ago, mammoths lumbered through what is now Kensington in central London — along with the first early humans to reach Britain. A portal to that distant world opens in mid-February at the Natural History Museum. The exhibition showcases landmark findings from the Ancient Human Occupation of Britain project, a collaboration between palaeontologists, archaeologists and geologists to craft a calendar of early human activity across the isles. Bringing together objects such as the world's oldest wooden spear and the skull of Britain's earliest known Neanderthal, the show will reveal a bigger picture — waves of humans arriving over the tumultuous millennia of the Pleistocene period, only to be beaten back repeatedly by glaciations.

## Pterosaurs: Flight in the Age of Dinosaurs

*American Museum of Natural History, New York*  
5 April 2014 – 5 January 2015

The most terrifying beasts (and the earliest vertebrates) ever to take to the air will throng a vast gallery in a temporary exhibition at



Pterosaurs such as *Tupandactylus* ruled the skies in the Cretaceous period.

directed by Wally Pfister, that techno-epiphany reportedly arrives when a computer scientist

uploads the brain of her assassinated husband, an artificial-intelligence researcher, into a computer. Does this brave new consciousness herald utopia or dystopia? Rebecca Hall, Johnny Depp and Morgan Freeman star.

#### Colour

National Gallery, London  
18 June – 7 September

From ochre to neon optics, colour has obsessed visual artists from prehistory on, although its maintenance has troubled conservators since at least the nineteenth century. In the National Gallery's 700-year overview of hue in paintings, glass, textiles and ceramics — which includes substantial input from the gallery's groundbreaking science department — the experimentation of colourists from the early Renaissance to the Impressionist era forms the base layer. The show explores the production of pigments, from the grinding of minerals to the formulation of acrylic polymers, as well as the challenges in rendering colours.

The trade routes that brought pigments

the American Museum of Natural History.

Pterosaurs, the flying reptiles that dominated the airspace from 220 million to 65 million years ago, will be explored through fossils, models and interactive displays, and contrasted with bats, birds and other animals that have evolved the remarkable ability to fly. This family-oriented show will celebrate these animals, which had wingspans that could — in the case of *Quetzalcoatlus northropi* — exceed 10 metres.

#### Transcendence

Director Wally Pfister  
Opens 18 April

The idea of the 'technological singularity' has been knocking around for decades, envisioned by mathematician John von Neumann and futurists including Ray Kurzweil as the moment when advances in artificial intelligence tip humanity into a radical new mode of being. In this much-anticipated science-fiction blockbuster

from caravan to canvas provide a fascinating historical context.

#### The Valley of Astonishment

Young Vic, London  
20 June – 12 July 2014

Theatrical legend Peter Brook has long been inspired by the wilder shores of neurology and mystical Islam. In *The Valley of Astonishment*, Brook and co-director Marie-Hélène Estienne mix the two. True stories of people with synaesthesia — a neurological condition in which the senses are mixed, so colours might be tasted or heard — are woven



into elements from the Sufi poet Attar's sublime twelfth-century epic *The Conference of the Birds*, from which the play's title derives. The formidable cast includes Kathryn Hunter and Marcello Magni.

#### From Atlantis to Today: Person, Nature, Catastrophes

Reiss-Engelhorn Museum, Mannheim, Germany  
7 September 2014 – 1 March 2015

Why do we mythologize catastrophes? This major exhibition takes as its unusual theme how different cultures have responded to natural disasters from antiquity until the present day. Simulations allow visitors to experience the sensations of the stranded as, for instance, Hurricane Katrina hammered and flooded New Orleans in 2005. And hundreds of objects related to disaster are on display, including a statue of Roman emperor Titus Augustus, who helped Pompeii's survivors after the AD 79 eruption of Vesuvius buried the town, and was then condemned by the populace for supposedly triggering the disaster.

#### Grand reopening

Wellcome Collection, London  
October

After a £17.5-million (US\$29-million) expansion and refurbishment, London's Wellcome Collection — a showcase for the links between medicine, art and daily life — will emerge radically recast and with several new spaces. A thematic gallery will host long-term exhibitions — the first taking on



A simulation of how the Imperial War Museum London's atrium will look when it opens in July.

#### First World War Galleries

Imperial War Museum London  
Opening July

To coincide with a huge programme of events commemorating the 100th anniversary of the start of the First World War, the Imperial War Museum London is set to open these themed galleries. Through interactive digital displays, audio and objects, visitors will explore the rapid escalation in industrial production that ensured that troops were fed and armed. The galleries will also depict a soldier's daily life, from psychological trauma to grappling with military technologies such as tanks and aeroplanes. The museum's refurbished atrium will display big hardware including a V2 rocket, Spitfire plane and T34 tank.

► the pioneers of sex research — and the current gallery will be spruced up before relaunching in October with a show on forensics. The glorious reading room will be made open to all as a place in which objects gathered by medical collector extraordinaire Henry Wellcome keep company with rare books, art and more.

### Interstellar

Director Christopher Nolan  
Opens 7 November

In a future near you, societal order has collapsed and the remnants of NASA are cobbled together to investigate a tear in the fabric of the Universe. *Interstellar*, already touted as one of the big films of 2014, will be a long-awaited cinematic outing for the ideas of theoretical physicist Kip Thorne, who advised on the venture. Breaking away from the idea that space exploration is limited to the Solar System, Thorne plays with the possibility of time travel using wormholes — ‘warps’ in space-time that serve as shortcuts to distant parts of the Universe. Christopher Nolan, who bent minds with 2010’s heist-within-a-dream-within-a-dream thriller *Inception*, directs.

### Russia’s Space Quest

Science Museum, London  
Autumn 2014

In 2014, Russia and Britain celebrate a joint year of culture, and the programme’s flagship event will be this showing of a remarkable collection of Soviet space artefacts. Visitors will be able to savour the sight of the capsules that carried cosmonauts aloft and the rocket engines that powered them, alongside smaller items from personal memorabilia to spacesuits. A collaboration with Moscow’s Memorial Museum of Cosmonautics and

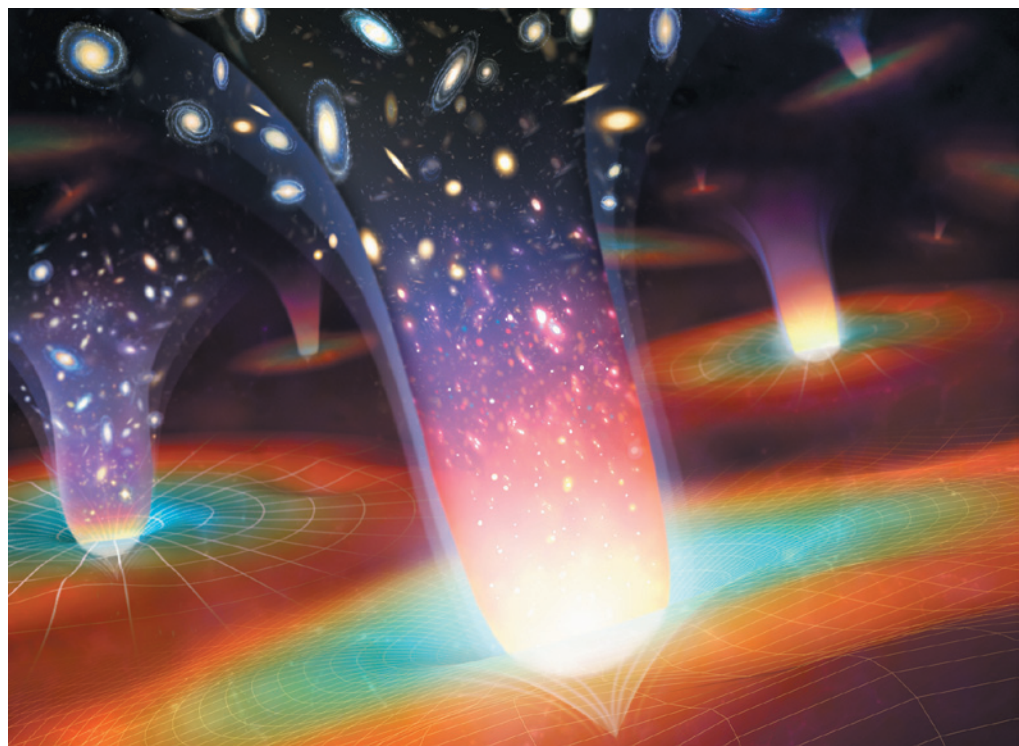
RYKOFF COLLECTION/CORBIS



Soviet propaganda celebrated the first human trip into space, by Yuri Gagarin.

the Russian space agency, Roscosmos, the show represents the most significant collection of such items ever permitted to leave Russia. ■

**Daniel Cressey** is a reporter for *Nature* in London. Additional reporting by Alison Abbott and Barbara Kiser.



### PHYSICS

## Chasing universes

Andrew Liddle contemplates an accomplished explication of the multiverse.

Having trouble understanding the Universe? Try this instead: imagine  $10^{500}$  possible universes, all different, and consider our place within this ensemble. Not randomly chosen, because our location should satisfy some basic conditions, such as habitability for intelligent species able to ask about their place in the cosmos. Can such a multiverse help us to fathom our Universe?

Cosmologist Max Tegmark has written an engaging and accessible book, *Our Mathematical Universe*, that grapples with this multiverse scenario. He aims initially at the scientifically literate public, but seeks to take us to — and, indeed, beyond — the frontiers of accepted knowledge. His explication of these ideas is more ambitious and individualistic than books on this topic by Leonard Susskind and Alex Vilenkin, for instance.

Multiverse theory stands in stark opposition to the belief that there should be some reason, perhaps a Theory of Everything, that determines physical laws such as the types

of particle that exist and the ways in which they interact. In the multiverse picture, it is all an accident.

### Our Mathematical Universe: My Quest for the Ultimate Nature of Reality

MAX TEGMARK  
Knopf: 2014.

What we know as ‘constants’ of nature, such as the strength of gravity or the proton-to-neutron mass ratio, happen to have particular values here, but in distant regions beyond our sight they may take other values and produce universes with very different properties — perhaps an absence of complex atoms and molecules, and hence of life.

Once seen as a fringe interest of dubious scientific validity, the multiverse has developed a serious following. Steven Weinberg used it in 1987 to predict that our observable Universe ought to have a non-zero cosmological constant, probably of a magnitude great enough to accommodate the acceleration of the Universe’s expansion. To everyone’s surprise, this was verified a decade later through observations of distant supernovae by two teams of astronomers. Those who led the work, Saul Perlmutter, Adam Riess and Brian Schmidt, won the 2011 Nobel Prize in Physics. Subsequently, string theory and inflationary cosmology were recognized as

➔ **NATURE.COM**  
For more on the  
multiverse, see:  
[go.nature.com/mqc2jd](http://go.nature.com/mqc2jd)



providing a setting that could predict, or at least motivate, the existence of a multiverse.

Tegmark's book captures two trends in contemporary science writing: scientific autobiography and the popular book as manifesto, expressing a body of personal scientific ideas ill-suited to traditional academic journals. Accordingly, Tegmark interweaves the science with stories of personal contributions to the endeavour. Fortunately, he is an engaging host. Tegmark makes his manifesto explicit by chopping his research life into two parts. Around a quarter of the book covers the 'sensible' work on constraining cosmological models from data. The rest is the outlandish part on quantum realities and multiple universes — even including an e-mail from a (sadly unnamed) senior academic advising him to desist before he destroys his career. It is clear where Tegmark's priorities lie: this book is his statement on the multiverse as a valid model for reality.

Tegmark likes the multiverse so much that he doesn't settle for just one; instead, he offers four different levels of multiverse. In the first, we simply have our own Universe, with its physical laws, extending forever. Shockingly, this is sufficient to ensure that, somewhere far away, there are exact replicas of you reading this review, on exact replicas of Earth. It might even be enough to imply that you are more likely to exist within a simulation of reality than in reality itself (whatever that means).

In the second incarnation, perhaps the most popular among proponents, physical laws vary within the multiverse so that distant regions can be considered to be distinct universes. This version is necessary to explain, for instance, the cosmological constant and other apparent coincidences in physical laws such as the stability of neutrons within nuclei. In the third level, the parallel universes may exist only as quantum mechanical states.

The culmination that Tegmark seeks to lead us to is the "Level IV multiverse". This level contends that the Universe is not just well described by mathematics, but, in fact, is mathematics. All possible mathematical structures have a physical existence, and collectively, give a multiverse that subsumes all others. Here, Tegmark is taking us well beyond accepted viewpoints, advocating his personal vision for explaining the Universe.

This is a valuable book, written in a deceptively simple style but not afraid to make significant demands on its readers, especially once the multiverse level gets turned up to four. It is impressive how far Tegmark can carry you until, like a cartoon character running off a cliff, you wonder whether there is anything holding you up. ■

**Andrew Liddle** is a theoretical cosmologist at the Institute for Astronomy, University of Edinburgh, UK.  
e-mail: [arl@roe.ac.uk](mailto:arl@roe.ac.uk)

## Books in brief



### The Accidental Universe: The World You Thought You Knew

Alan Lightman PANTHEON (2014)

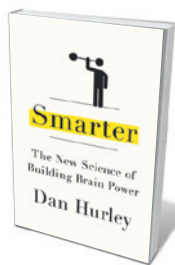
Theoretical physicist Alan Lightman's meditation on how recent findings in science shape our concept of self and Universe unfolds with the mesmeric calm of a vessel in space. That is, until he treats us to some split-second encounter with a sliver of the totality — such as the eye of a flying osprey. In seven elegant essays on aspects of the Universe, Lightman takes us from the idea of an accidental cosmos, prompted by multiverse theory, to the Higgs boson, digital disembodiment and the cosmic evanescence that fits so poorly with our longing for permanence.



### The Science of Cheese

Michael Tunick OXFORD UNIVERSITY PRESS (2013)

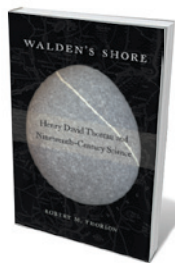
From "smear-ripened" Swiss tilsit to the maggot-riddled casu marzu of Italy, cheese can carry a whiff of the surreal. Chemist Michael Tunick tours a sample of the 2,000 known varieties, mingling science (biology, chemistry, physics, nutrition and climatology) and cultural lore to make an accessible whole. If you have ever wondered what links Limburger with foot perspiration (answer: short-chain fatty acids), or how to make mozzarella at home, Tunick is your man. And the world's most expensive cheese? Made from moose milk on a Swedish farm, it will set you back US\$1,000 a kilogram.



### Smarter: The New Science of Building Brain Power

Dan Hurley HUDSON STREET PRESS (2013)

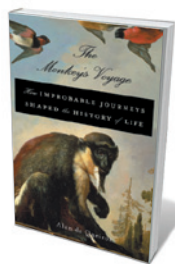
Is "fluid" intelligence — the ability to think on your feet and discern patterns — teachable? In this trip through the findings on and controversies around brain training, science journalist Dan Hurley proves an able, often caustically humorous guide. He starts with 2008 research on working-memory training by psychologists Susanne Jaeggi and Martin Buschkuhl, then trawls research in areas such as gaming and visual attention processing. After taking the pulse at science conferences and turning guinea pig to test a range of techniques, Hurley admits to cautious optimism.



### Walden's Shore: Henry David Thoreau and Nineteenth-Century Science

Robert M. Thorson HARVARD UNIVERSITY PRESS (2013)

In his 1854 masterpiece *Walden*, the US writer and naturalist Henry David Thoreau invites us to "wedge our feet downward ... till we come to a hard bottom and rocks in place, which we can call reality". Geologist Robert Thorson obliges, focusing on Thoreau as a flinty amateur geologist to reinject science into his literary legacy. Thoreau, Thorson persuasively argues, was as grounded in rock as he was in the elemental understanding of the cosmos sought by the Transcendentalist movement.



### The Monkey's Voyage: How Improbable Journeys Shaped the History of Life

Alan de Queiroz BASIC BOOKS (2014)

Biogeography is undergoing a sea change, argues Alan de Queiroz. The dominant theory of global species dispersal previously centred on the break-up of the supercontinent Gondwana, starting some 160 million years ago. Now, the idea of species traversing oceans is gaining ground. Perhaps the most compelling scenario is the 'monkey transfer' from Africa to South America, envisioned as a simian troop hitching a ride on an uprooted, floating tree. [Barbara Kiser](#)

# Correspondence

## Targeted MOOC captivates students

In our experience, massive open online courses (MOOCs) can reach non-traditional and disadvantaged learners if they address a recognized need, support the educational requirements of the intended cohort and enable learning with tangible outcomes (see E. J. Emanuel *et al.* *Nature* **503**, 342; 2013).

The Wicking Dementia Research and Education Centre at Australia's University of Tasmania has developed a MOOC on understanding dementia. This attracted almost 10,000 people from more than 60 countries in July, of which 89% were women, 70% were over the age of 40, and only 17% were educated beyond a bachelor's degree (compared with 44% reported by Emanuel and colleagues).

The course is tailored to the educational needs of the care workforce and family-based carers who support the more than 44 million people with dementia worldwide.

Our cohort-centric approach involved structuring the course to support non-traditional learners, including providing online technical and teaching support.

The completion rate for the 11-week course was 39%, which is considerably better than the international average for MOOCs (see C. Parr *Times Higher Education* 9 May 2013; [go.nature.com/p25g67](http://go.nature.com/p25g67)).

**Carolyn King, Andrew Robinson, James Vickers**  
*Wicking Dementia Research and Education Centre, University of Tasmania, Hobart, Australia.*  
[carolyn.king@utas.edu.au](mailto:carolyn.king@utas.edu.au)

## Storm-surge impact depends on setting

A storm surge on 5–6 December threatened urban centres and rural communities around the southern North Sea in a similar way to such an event 60 years ago. Causing more than 2,000 deaths, the 1953 flood was western Europe's most devastating in 100 years in terms of loss of life. Last month, however, a disaster was averted by advances in storm-surge forecasting, improved defences, early-warning systems and integrated crisis management.

Immediately after the surge, we made high-resolution measurements of maximum water levels, focusing on obvious debris lines, erosion points on earthen bank defences

and water marks on buildings along the 45-kilometre-long northern coastline of Norfolk in the United Kingdom. These confirm that flood levels were similar to, and in places exceeded, those in the 1953 disaster.

There was considerable variation in the mean height of peak water levels along the shore (the maximum difference between measurement stations was more than 1.2 metres). This reflects the combined effects of tide, surge and wave run-up, which has a strong local component. For this coastline of barrier islands, spits and tidal embayments, these observations indicate that the coastal setting and extent of coastal ecosystems (such as mudflats and salt marshes) are critical in determining the pattern of storm-surge impacts.

Such differences become crucial when properties, infrastructure and lives are threatened by sea flooding (see also J. D. Woodruff *et al.* *Nature* **504**, 44–52; 2013). These factors should be incorporated into hydrodynamic modelling and forecasting efforts, to help fine-tune early-warning systems and evacuation planning.  
**Thomas Spencer** *University of Cambridge, UK.*

[ts111@cam.ac.uk](mailto:ts111@cam.ac.uk)

**Susan M. Brooks** *Birkbeck, University of London, UK.*  
**Iris Möller** *Cambridge Coastal Research Unit, UK.*

## Whistle-blowers have a tough time

Whistle-blower cases that go on forever are not uncommon (see *Nature* **503**, 454–457; 2013). The cold conclusion is that the whistle-blower may survive, but the odds are against him or her.

I have worked with whistle-blowers for more than 35 years as an expert witness in court cases and as author of the forthcoming book *Don't Kill the Messenger* (see [www.whistleblowing.us](http://www.whistleblowing.us)), and find that they are hard to silence. The truth-telling part of their brain seems to override the health and safety part, so they will endure all forms of retaliation for the sake of truth.

Institutions can also be very slow to admit to any mistakes on their watch. This factor delays adjudication and makes it harder for the whistle-blower to prove anything in court.

**Don Soeken** *Whistleblower Support Fund, Ellicott City, Maryland, USA.*  
[helpline@tidalwave.net](mailto:helpline@tidalwave.net)



## NATURE'S READERS COMMENT ONLINE

*A sample of responses from the debate on the reproducibility drive (M. Bissell *Nature* **503**, 333–334; 2013).*

### Nitin Gandhi says:

The very fact that we have to take the issue of replication so seriously and spend lots of time and money on it in these hard times speaks out loudly that things are not right in biomedical research.

### William Gunn says:

The Reproducibility Initiative aims to make science work better for everyone [see [go.nature.com/v5c1js](http://go.nature.com/v5c1js)]. The worst that could happen is that we learn a lot about what level of reproducibility to expect and how to reliably build on a published finding. At best,

funders will start tacking a few per cent on to grants for replication purposes and publishers will start asking for it. That can only be good for science as a whole.

### Anonymous says:

I would be a rich man if I had received a penny for every time I heard the expression “in our hands” at a scientific lecture. I disagree that “the push to replicate findings could shelve promising research and unfairly damage the reputations of careful, meticulous scientists”. I believe that the opposite is true.

Scientists should be encouraged to report and publish when they fail to replicate each other's experiments. That will help science (but maybe not scientific careers) progress much faster.

### Irakli Loladze says:

The current system does not penalize for publishing sexy but non-reproducible findings. In fact, such publications boost the chances of getting another grant. It is about time to end this vicious cycle that benefits a few but hurts science at large.



# Frederick Sanger

## (1918–2013)

Double Nobel–prizewinning genomics pioneer.

**F**rederick Sanger, ‘the father of genomics’, was one of just four scientists to win two Nobel prizes and the only one to receive both in chemistry. Both were awarded for the invention of methods to determine the order of the biological building blocks of life.

Sanger will be remembered especially for developing techniques to read out the As, Cs, Gs and Ts in a strand of DNA. This work provided the means to decipher genetic material and led to his second prize, which he shared with Paul Berg and Walter Gilbert in 1980. In the 1990s, Sanger’s eponymous method was used by laboratories around the world to work out the sequence of the human genome.

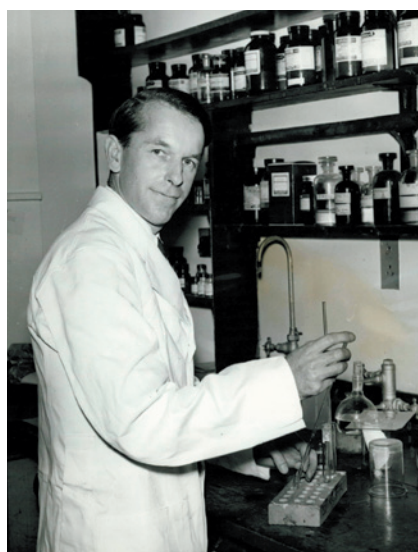
His first prize came in 1958 for his discovery of how amino acids are strung together in the protein insulin. In the 1950s, many thought that the amino acids within a protein were arranged randomly, but Sanger proved beyond doubt that they instead form a unique sequence. Although he made light of this conclusion, saying that those who knew about proteins expected this outcome, the knowledge that proteins had a precise sequence suggested that this information must be codified in DNA.

Sanger, who died in Cambridge, UK, on 19 November aged 95, was born in 1918 in Gloucestershire. Raised as a Quaker, he learned self-reliance and practical manual skills as a schoolboy. These aptitudes were used to great effect in his laboratory and in building sailing boats.

He developed an interest in science from his physician father and his older brother, with whom he enjoyed the outdoors. In 1939, he graduated in biochemistry from St John’s College, Cambridge. A conscientious objector, he stayed on at the University of Cambridge during the Second World War to study the nutritional benefit of lysine in potatoes under biochemist Albert Neuberger. In 1940, Sanger married Margaret Joan Howe, an economics graduate. They had three children and remained married until her death in 2012. Sanger ascribed his wife and his fellow researchers key roles in his success.

After receiving his PhD in 1943, Sanger began the research that led to his first Nobel prize, working out how amino acids link up in the two polypeptide chains of insulin. He labelled the ends of the separate chains with a yellow dye, then hydrolysed them to amino acids and identified the tagged amino acid in each case. After using acid

and enzymes to split each chain into defined fragments, he tagged purified fragments with the dye and repeated the process. From this, and from the amino-acid composition of the fragments, he deduced the order of



amino acids in the intact protein, rather like building up a picture from the pieces of a jigsaw puzzle.

Sanger preferred to be in the background but was not afraid to use his clout. He supported a successful bid to the UK Medical Research Council (MRC) to build the Laboratory of Molecular Biology in Cambridge, which opened in 1962. Here, Sanger spent the rest of his active scientific life.

After first working out ways to sequence RNA molecules, by which sequence information in genes is transferred into the sequences of proteins, Sanger took up the challenge of sequencing genes themselves. He developed a method that used enzymes to copy fragments of DNA. Four reactions were set up side by side, each supplied with the four standard building blocks, or nucleotides, (As, Cs, Gs and Ts), one of which was labelled with radioactive atoms. Each reaction also contained a modified version of A, C, G or T. Unlike standard nucleotides, these ‘chain terminators’ did not allow the DNA strand to grow further after they had been incorporated. Interrupted copies were separated according to their size on gels by an electric current and exposed to photographic film, allowing the radioactivity to produce the now-iconic ‘ladders’ of dark

bands. These bands revealed the length of the DNA copy and allowed the sequence to be read simply. By combining sequences of many DNA fragments, the sequence of the larger DNA molecule from which the fragments were derived could be deduced.

Sanger demonstrated the power of his method by sequencing genomes of ever-increasing size, starting with a simple bacterial virus (5,386 nucleotides) in 1977, then the DNA in the mitochondria of human cells (16,569 nucleotides) in 1981 and, finally, the genome of a complex bacterial virus, bacteriophage lambda (48,502 nucleotides), in 1982.

In 1993, nine years after Sanger retired, the Wellcome Trust and the MRC opened the Sanger Centre (now the Wellcome Trust Sanger Institute) near Cambridge, where a considerable part of the human genome was decoded with the technique he developed. In the 2000s, Sanger sequencing gradually gave way to faster, cheaper techniques that detect nucleotides as they attach to growing DNA strands. But Sanger sequencing remains the gold standard. The highly accurate technique is increasingly being applied to the genomes of individual humans and even individual cells within tumours. Sanger’s impact on biology is as dramatic as that of Charles Darwin.

Sanger was happiest at the laboratory bench, where he worked tirelessly and single-mindedly. He performed elegant experiments with simple apparatus to solve extremely difficult problems. In so doing, he inspired younger scientists and attracted some of the best biologists in the world to Cambridge.

Sanger was famously understated, but he knew that he was an extraordinary scientist, and when the occasion demanded it he was prepared to say so. When colleagues assembled after the announcement of his second Nobel Prize, one praised his characteristic modesty. Sanger responded: “I want you all to know that I think that I am bloody good.” He was showered with awards and quietly enjoyed the recognition. After retirement, he continued to build boats and developed a magnificent English garden. ■

**John Walker** received the Nobel Prize in Chemistry in 1997. From 1974 to 1984, he worked alongside Frederick Sanger at the Medical Research Council Laboratory of Molecular Biology.  
e-mail: [walker@mrc-mbu.cam.ac.uk](mailto:walker@mrc-mbu.cam.ac.uk)

## MALARIA

# Resistance nailed

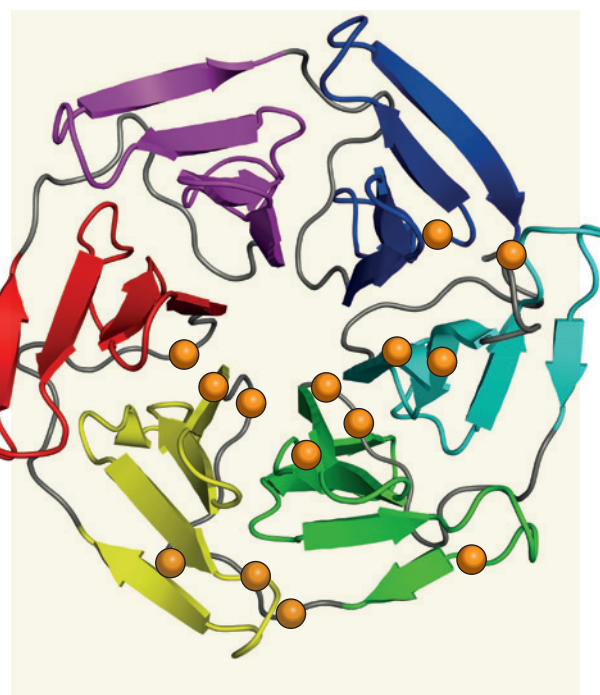
A series of *in vitro*, genomic, ecological and epidemiological studies has pinpointed gene mutations in the malaria parasite *Plasmodium falciparum* that play a key part in resistance to artemisinin-based antimalarial drugs. [SEE ARTICLE P.50](#)

CHRISTOPHER V. PLOWE

A fragile consensus that the global eradication of malaria is possible, and prospects for eventually achieving this audacious goal, are being threatened by the emergence in south-east Asia of parasite resistance to the drug artemisinin and its derivatives. On page 50 of this issue, Ariei *et al.*<sup>1</sup> report welcome news: they have identified a molecular marker of artemisinin-resistant malaria that can be used to map resistance and guide efforts to eliminate it.

Artemisinin-based combination treatments have contributed to reductions in the global burden of malaria, prompting the Bill & Melinda Gates Foundation and the World Health Organization to issue a call in 2007 for an international push towards malaria eradication<sup>2</sup>. Artemisinin drugs normally clear malaria parasites from the blood of a patient within two days of starting treatment; now, however, increasing numbers of *Plasmodium falciparum* infections in western Cambodia, southern Vietnam, eastern Myanmar and western Thailand take up to five days to clear. In some areas, artemisinin-based combination therapies are starting to fail completely, with persistence of both infection and clinical illness after what should be curative treatment.

Efforts to contain artemisinin resistance in southeast Asia and to eliminate malaria would be aided enormously by the identification of a molecular marker for this drug resistance. Such markers are available for resistance to other antimalarial drugs for which the genetic determinants of resistance in the parasite are known. However, neither the mechanism of artemisinin action nor the mechanism(s) of resistance are understood. Examinations of the *P. falciparum* genome for regions of recent strong evolutionary selection, and targeted and genome-wide association studies, have implicated two adjacent regions on chromosome 13 as potential sites of a resistance-determining gene or genes<sup>3,4</sup>. Through dogged determination and a remarkable combination



**Figure 1 | Propeller mutations.** Ariei *et al.*<sup>1</sup> show that mutations in a *Plasmodium falciparum* gene that encodes the kelch protein K13 are associated with both *in vitro* and clinical measures of artemisinin-resistant *P. falciparum* malaria in Cambodia. The mutations (orange spheres) encode amino-acid changes in the 'propeller blades' of this protein, which resembles a child's pinwheel and is thought to be involved in various protein-protein interactions.

of approaches, Ariei and colleagues seem to have won the race to identify if not the gene, at the very least a key gene, responsible for artemisinin resistance.

In what seemed like a long shot, the researchers laboriously grew an artemisinin-sensitive parasite isolated from a Tanzanian individual in culture for five years, exposing it intermittently to artemisinin. The drug was removed when the parasites' growth faltered, and replaced after they bounced back. After 60 cycles of drug pressure, the proportion of parasites surviving a pulse of artemisinin had increased from less than 0.01% to more than 10%. Genome sequencing of this population revealed eight single-nucleotide mutations (single nucleotide polymorphisms, or SNPs) in seven genes that were present in the resistant parasites but not in their

siblings grown in a parallel culture without drug exposure.

The suspects were cornered. The culprit was then identified when the authors looked for these candidate mutations in parasite lines from Cambodia that had varying susceptibilities to artemisinin drugs *in vitro*<sup>5</sup>. After ruling out candidate genes from the Tanzanian isolate that showed no sequence variation in the Cambodian isolates, and genes containing SNPs that were not associated with *in vitro* resistance, a single gene remained with resistance-associated SNPs. The gene is located on chromosome 13, within a candidate region identified in a recent genome-wide association study of clinical artemisinin resistance<sup>4</sup>.

The gene in question encodes a kelch protein called K13. Kelch proteins are involved in a variety of protein-protein interactions, and contain several regions of repeating amino-acid sequences, each forming a 'propeller blade' (Fig. 1). Further evidence for the central role of K13-propeller SNPs in resistance was provided by Ariei and colleagues' ecological survey of malaria parasites from several Cambodian provinces. K13-propeller mutations were rare or absent in sam-

ples from provinces with virtually no documented resistance, but were widespread in provinces where resistance has been reported. Moreover, their prevalence in these provinces has increased during the past decade, contemporaneously with increases in the prevalence of resistance.

The authors went on to show that K13 SNPs were also strongly correlated with delayed parasite clearance following artemisinin treatment in clinical trials. And a final piece of evidence came from subpopulations of Cambodian parasites that can be segregated into sensitive and resistant groups<sup>6</sup>. The prevalence of K13 SNPs not only correlated with resistance among these subpopulations, they actually did a better job of explaining resistance than did population groupings: sensitive parasites assigned on the basis of their



genomic profile to 'resistant' subpopulations had wild-type K13, and resistant parasites belonging to 'sensitive' subpopulations carried the SNPs.

Definitive proof that K13-propeller mutations confer artemisinin resistance will come from genetic transformation of drug-sensitive parasites into resistant ones by the replacement of their wild-type K13 gene with a mutated gene. It is of course possible — even probable — that other genes contribute to artemisinin resistance, but this study leaves little doubt that K13 is a major determinant of resistance to these drugs in *P. falciparum* malaria. Further study of gene variants in the chromosomal vicinity of the K13 gene will ascertain whether resistance arose once in western Cambodia and then spread — in which case the genomic regions containing the K13 SNPs will show extended surrounding sequence similarity, indicating common ancestry — or whether it emerged independently in different

geographical locales. If resistance has arisen independently in many areas, local containment efforts will be futile and only regional elimination offers any hope of preventing its spread to Africa, where the arrival of drug-resistant Asian parasites has previously led to marked increases in malaria hospitalizations and deaths<sup>7</sup>.

Validation of this molecular marker of artemisinin resistance outside Cambodia will be easily achieved, and mapping of the marker throughout southeast Asia is already under way, thanks to early sharing of the results of this study with local malaria-control workers and researchers. With at least 17 SNPs residing in the propeller domains of K13, only one of which is found in any one parasite, sequencing of the K13 gene will initially be necessary to map resistance. But if a few specific SNPs emerge as being predictive of resistance in different settings, switching to rapid molecular assays using DNA extracted from dried blood

spots will accelerate translation of this research finding into a practical tool for public-health surveillance. ■

**Christopher V. Plowe** is at the Howard Hughes Medical Institute and the Center for Vaccine Development, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA.

e-mail: [cplowe@medicine.umaryland.edu](mailto:cplowe@medicine.umaryland.edu)

1. Arie, F. *et al.* *Nature* **505**, 50–55 (2014).
2. Roberts, L. & Enserink, M. *Science* **318**, 1544–1545 (2007).
3. Cheeseman, I. H. *et al.* *Science* **336**, 79–82 (2012).
4. Takala-Harrison, S. *et al.* *Proc. Natl Acad. Sci. USA* **110**, 240–245 (2013).
5. Witkowski, B. *et al.* *Lancet Infect. Dis.* **13**, 1043–1049 (2013).
6. Miotto, O. *et al.* *Nature Genet.* **45**, 648–655 (2013).
7. Trape, J.-F. *et al.* *C.R. Acad. Sci. III* **321**, 689–697 (1998).

This article was published online on 18 December 2013.

## EXTRASOLAR PLANETS

# Cloudy with a chance of dustballs

**The flat and featureless transmission spectra of two intermediate-sized extrasolar planets, observed during the planets' passage across their host stars, shed light on the properties of their atmospheres. SEE LETTERS P.66 & P.69**

JULIANNE MOSES

Flat spectra do not typically excite astronomers, but there are times when a lack of spectral features tells you something interesting. Such is the case with observations of two separate sub-Jupiter-sized extrasolar planets made using the Hubble Space Telescope and reported by Knutson *et al.*<sup>1</sup> and Kreidberg *et al.*<sup>2</sup> on pages 66 and 69 of this issue, respectively.

The first planet, GJ 436b, has a mass and radius slightly greater than Neptune's. The second, GJ 1214b, is smaller, with a radius roughly 2.7 times that of Earth. Both exoplanets orbit very close to their host stars and are therefore quite warm by Earth standards. Owing to intensive observational scrutiny since their respective discoveries, these two planets have become the archetypes of the new 'Neptune-class' and 'super-Earth' categories of exoplanets.

Although the first extrasolar planets discovered — in the 1990s and 2000s — tended to be hot, massive, hydrogen-dominated worlds that most closely resemble Jupiter, astronomers have been methodically and efficiently chipping away at the harder-to-observe regime of smaller, denser and cooler Earth-like planets.

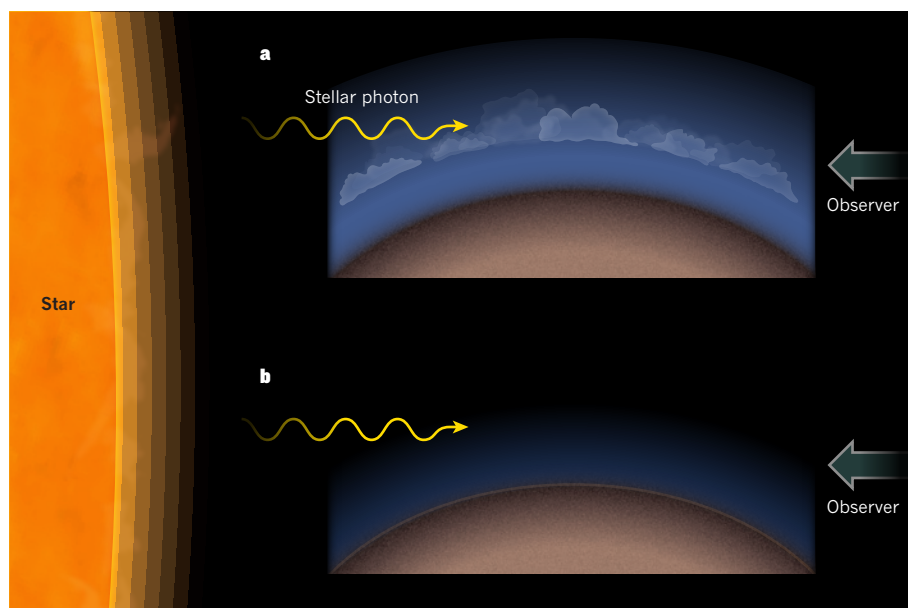
Recent ground- and space-based surveys demonstrate<sup>3–5</sup> that planets of sizes ranging between those of Earth and Neptune overwhelmingly dominate the observed exoplanet population. But what are these intermediate-sized exoplanets really like? Our Solar System does not provide sufficient clues about these planets, because we have only Earth and Venus at one end of the scale and cold Uranus and Neptune at the other to serve as examples. Do these mid-sized exoplanets have rocky surfaces, like Earth and Venus? Are they fluid planets with thick, deep atmospheres relatively rich in hydrogen and volatile elements, like Uranus and Neptune? Are they 'water worlds' with steam atmospheres overlying deep oceans? Are the atmospheres of these planets thick or thin, consisting predominantly of hydrogen, water, carbon dioxide or nitrogen, or something more exotic? Observations of GJ 436b and GJ 1214b provide important hints.

The orbital planes of both GJ 436b and GJ 1214b are almost exactly edge-on as seen from Earth, so that the planets periodically transit — pass directly in front of — their host stars, causing slight dips in the amount of light seen from the system. The depth of these transit dips allows the planetary radius to be

determined, and their wavelength dependence provides information about atmospheric composition. During transits, the stellar light passes through the planet's atmosphere on its way to the observer. For a planet with an extended atmosphere, the apparent size of the planet can vary with observed wavelength because more stellar light is blocked at high altitudes at wavelengths for which atmospheric constituents have strong absorption bands. Conversely, more light passes through at low altitudes at wavelengths for which atmospheric constituents are less absorbing.

GJ 1214b is known to have a relatively flat (wavelength-independent) transmission spectrum, with weak absorption features at best<sup>6,7</sup>. Because the bulk density of the planet suggests that it must have a gaseous envelope, the two leading theories for explaining the flat transmission spectrum involve either widespread, high-altitude clouds or a hydrogen-poor atmosphere dominated by a high-molecular-weight constituent such as water or carbon dioxide (Fig. 1). In both cases, the stellar light at most wavelengths would be extinguished fairly abruptly within a small vertical region of the atmosphere.

In their study, Kreidberg and colleagues present near-infrared transmission spectra for GJ 1214b that finally allow one of the two competing theories to be ruled out for this super-Earth exoplanet. The extremely precise spectra, obtained from the Wide Field Camera 3 (WFC3) on board the Hubble Space Telescope, demonstrate that GJ 1214b's transmission spectrum is so flat and featureless in a wavelength region between about 1.1 and 1.6 micrometres that high-altitude clouds provide the only plausible explanation. The observations are precise enough that spectral features from a cloud-free atmosphere dominated by heavy molecules such as water,



**Figure 1 | Exoplanetary atmospheres.** Flat transmission spectra of exoplanets during transit, such as those reported by Knutson *et al.*<sup>1</sup> and Kreidberg *et al.*<sup>2</sup>, can result from a planet with an atmosphere that either contains high clouds (a) or that is hydrogen poor with a high mean molecular weight (b). In the cloudy case, photons from the planet's host star are blocked abruptly when they encounter the cloud layer on their way to an observer on Earth. In the hydrogen-poor case, the high molecular weight of the atmosphere allows it to be bound tightly by gravity and therefore be vertically compressed, with large changes in density over relatively small vertical scales providing a relatively sudden absorption of all the stellar photons. If the planet had a clear (cloud-free), low-mean-molecular-weight atmosphere (not shown), atmospheric absorption features would be more prominent in the transmission spectrum.

methane, carbon monoxide or carbon dioxide would have been detectable if such an atmosphere were present on GJ 1214b. Even an atmosphere composed of 99.9% spectrally neutral nitrogen with 0.1% water can be rejected on the basis of the lack of water-absorption features.

Meanwhile, new WFC3 observations of GJ 436b presented by Knutson and colleagues point to a similarly flat and featureless transmission spectrum between 1.1 and 1.6  $\mu\text{m}$  for this Neptune-class planet. Given that one might expect the more massive GJ 436b to contain more hydrogen than GJ 1214b, the flat spectrum is, in this case, an even bigger surprise — a hydrogen-rich atmosphere would be vertically extensive, and expected trace species such as water and methane would have prominent deep absorption bands. However, unlike the situation for GJ 1214b, Knutson *et al.* demonstrate that a hydrogen-poor atmosphere (with or without clouds) and a hydrogen-rich atmosphere with high clouds are both statistically viable solutions to explain the observed flat transmission spectrum for GJ 436b. To distinguish between these scenarios, more precise moderate-resolution spectral observations at near-infrared wavelengths will be needed to unambiguously reveal any spectral features. Longer-wavelength eclipse observations<sup>8</sup>, acquired when the planet passes behind the star, could also help to discriminate between the two hypotheses.

Evidence is mounting that the hydrogen

fraction within a planet is a strong function of planet size<sup>9</sup>, so it is not necessarily an 'either-or' situation for explaining the flat transmission spectra of GJ 436b and GJ 1214b: the atmospheres could be cloudy and have a large mean molecular weight. However, high-altitude clouds on these two exoplanets would not resemble the clouds we see in the Solar System. Possible candidates include potassium

chloride or zinc sulphide 'dust' clouds. For the case of a relatively hydrogen-poor atmosphere, these two components would form clouds that are optically thick (opaque) enough at high altitudes on both planets that the transmitted stellar light would be abruptly blocked, leading to a flat transmission spectrum<sup>10</sup>. Alternatively, thick hazes such as those seen around Saturn's moon Titan could be produced from photochemical processing of atmospheric gases by ultraviolet stellar photons, although the lack of evidence for methane on either of these two planets<sup>2,8</sup> suggests that any photochemical hazes present would be decidedly different from those on Titan.

Hydrogen-poor or not, dust-shrouded or not, super-Earth and Neptune-class planets collectively represent an intriguing and populous type of extrasolar planet whose exotic atmospheres may have no true analogues in the Solar System. The transmission spectra presented here — flat and featureless, and yet full of information — provide one piece of the puzzle needed to characterize such planets. ■

Julianne Moses is at the Space Science Institute, Boulder, Colorado 80301, USA.  
e-mail: jmoses@spacescience.org

1. Knutson, H. A., Benneke, B., Deming, D. & Homeier, D. *Nature* **505**, 66–68 (2014).
2. Kreidberg, L. *et al. Nature* **505**, 69–72 (2014).
3. Mayor, M. *et al.* Preprint at <http://arxiv.org/abs/1109.2497> (2011).
4. Cassan, A. *et al. Nature* **481**, 167–169 (2012).
5. Batalha, N. M. *et al. Astrophys. J. Suppl. Ser.* **204**, 24 (2013).
6. Bean, J. L., Miller-Ricci Kempton, E. & Homeier, D. *Nature* **468**, 669–672 (2010).
7. Berta, Z. K. *et al. Astrophys. J.* **747**, 35 (2012).
8. Stevenson, K. B. *et al. Nature* **464**, 1161–1164 (2010).
9. Lopez, E. D. & Fortney, J. J. Preprint at <http://arxiv.org/abs/1311.0329> (2013).
10. Morley, C. V. *et al. Astrophys. J.* **775**, 33 (2013).

#### ARCHAIC HUMANS

## Four makes a party

**Adding the first high-quality Neanderthal sequence to genomic comparisons of archaic and modern humans sheds light on gene flow, population structure and adaptation, and suggests the existence of an unknown group. SEE ARTICLE P.43**

EWAN BIRNEY & JONATHAN K. PRITCHARD

Archaic humans have captured the popular imagination since the nineteenth century, when the remains of Neanderthals were discovered in the Neander valley of northern Germany and elsewhere in Europe. Until recently, Neanderthals and other archaic humans were known only from bones and various artefacts, but DNA-sequencing technology is now providing us with new perspectives on these early groups

and their relationships to modern humans. In this issue, Prüfer *et al.*<sup>1</sup> (page 43) report the first high-quality genome sequence of a Neanderthal individual. Their work adds to an emerging story about a tangled web of gene flow among modern humans and different early hominins (humans and archaic groups that are more closely related to humans than to chimpanzees), and hints at the existence of an unknown, highly diverged hominin group that contributed to this archaic gene pool.

Neanderthals are thought to have persisted

in southern Europe until around 30,000 years ago<sup>2</sup>, thus potentially overlapping with modern humans. As a result, there has long been interest in whether Neanderthals might have interbred with early Europeans. In the 1990s, the first comparisons of DNA sequences from modern humans and Neanderthals<sup>3,4</sup> suggested a rather simple story: that modern humans emerged from Africa during the past 100,000 years, and spread around the globe without receiving genetic contributions from hominins that had left Africa much earlier.

These early studies were based on sequences from mitochondrial DNA, which is easier than nuclear DNA to capture in ancient samples but represents only a tiny fraction of the human genome. However, the past few years have seen a revolution in our ability to obtain nuclear-genome sequences from ancient samples<sup>5–9</sup>, and these data are providing startling insights. One surprise was the first clear evidence for interbreeding between Neanderthals and modern humans<sup>5</sup>; another was the discovery of a second type of archaic hominin in Eurasia in addition to Neanderthals. This group, dubbed the Denisovans, is known mainly from the genome sequence of a single finger bone found in a cave in the Altai Mountains in Siberia<sup>6,7</sup>.

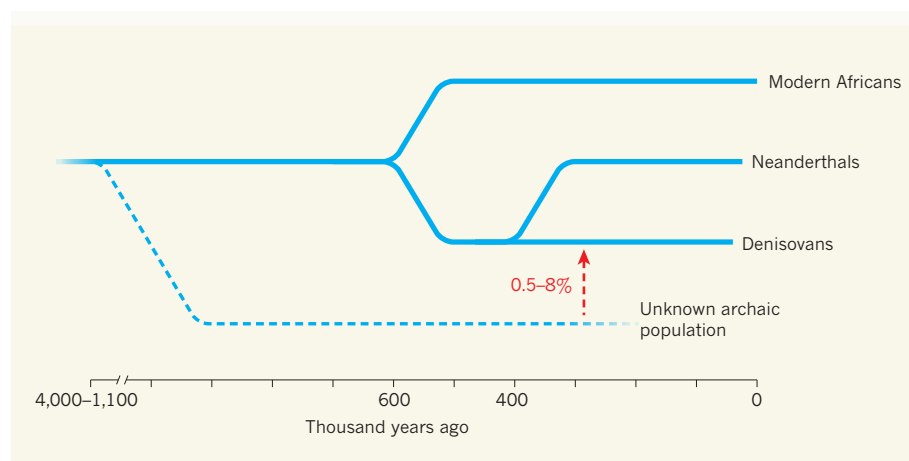
Although the Neanderthal bone from which Prüfer *et al.* derived their genomic sequence was found in the same Siberian cave, its owner is estimated to have lived several thousand years earlier than the Denisovan individual, and the two populations that the individuals represent are not closely related. The ancestors of Neanderthals and Denisovans diverged from the main human lineage about 600,000 years ago, and then split from each other around 400,000 years ago (Prüfer *et al.*

discuss these estimates and associated caveats in detail). Thus, Neanderthals and Denisovans were quite distinct populations, having been separated for roughly three times longer than any modern human populations.

Prüfer and colleagues' sequence comparisons provide further detail about the extent of interbreeding between the different hominin groups living during the Pleistocene period (see Fig. 8 of the paper<sup>1</sup>). The authors offer a more confident estimate of the Neanderthal contribution to the genomes of modern humans: about 2% for non-Africans (Africans have no detectable Neanderthal ancestry). They also report gene flow from Neanderthals into Denisovans that includes input at functionally important genomic regions involved in immunity and sperm function. Earlier work had shown that the main Denisovan contribution to modern humans is found in some populations in Oceania and, to a lesser extent, in east Asians<sup>6,7</sup>.

Most provocatively, Prüfer *et al.* find evidence for modest levels of gene flow into Denisovans of sequence that is different from that of any known group, implying that there is at least one more, so far undiscovered, archaic-hominin group (Fig. 1). Low levels of gene flow have been observed in other radiations of species, so evidence for inter-hominin breeding should not be a tremendous surprise<sup>10</sup>; however, it does seem that Eurasia during the Late Pleistocene was an interesting place to be a hominin, with individuals of at least four quite diverged groups living, meeting and occasionally having sex.

The Neanderthal and Denisovan genomes also share another intriguing feature: they both have extremely low genetic diversity, with only about two heterozygous sites



**Figure 1 | Gene flow from an unknown ancient population.** Prüfer *et al.*<sup>1</sup> calculate that modern Africans show greater genomic similarity to Neanderthals than to Denisovans. Average sequence divergence along the lineage leading to modern Africans is 7.47% since the last common ancestor with Neanderthals, and 7.71% since the last common ancestor with Denisovans (both numbers represent the percentage of divergence since the human–chimpanzee split). This difference is highly significant, and is inconsistent with a simple model in which the entire Neanderthal and Denisovan genomes come from the same source population. The best alternative model identified by the authors is that there was flow of a small contribution of genomic material (0.5%–8%) into Denisovans from a highly diverged, unknown population.



## 50 Years Ago

The Continental Shelf Bill, which received its second reading in the House of Lords on December 3, originated in the Conference on the Law of Sea at Geneva in 1958, which resulted in the Continental Shelf Convention and the High Seas Convention. The former, which the Government intends to ratify if the Bill becomes law, clarified international law concerning those large submarine areas outside the territorial seas where the depth of the water allows the natural resources of the sea-bed and subsoil to be exploited ... In the North Sea ... Britain will have rights over any deposits up to a line half-way across to Holland, Belgium and other coastal States, subject to any adjustments resulting from the negotiations that the Government proposes to undertake after ratifying the Convention.

From *Nature* 4 January 1964

## 100 Years Ago

Major H. G. Joly De Lotbinière has contributed to *The Quarterly Review* for October a valuable and timely article on the position of forestry in England and abroad, in which he reviews the principal timber resources of the world, and the steps that have been taken in England and elsewhere to provide for the future. As he points out, experts in every country are agreed that the world's supply of timber is rapidly diminishing, and that unless vigorous steps are taken in the afforestation of suitable waste lands a shortage of material must be experienced long before the close of the present century. The author indicates in a general way the lines on which the work of afforesting the sixteen million acres of mountainous and heath land in this country should be proceeded with, and urges the necessity for immediate action.

From *Nature* 1 January 1914



(sequence differences between the paired homologous chromosomes) per 10,000 nucleotides. This equates to only around one-quarter of the genetic diversity of modern humans. The Neanderthal individual sequenced by Prüfer *et al.* had reduced heterozygosity in part because she was inbred (her parents were as related as half-siblings). However, the authors' analysis suggests that the primary cause of the low variability is that both groups had extremely small effective population sizes for the preceding 100,000 years or more.

Not only are these diversity estimates low compared with the genetic diversity of modern humans, they are also among the lowest levels of genetic diversity reported for any organism<sup>11</sup>. These small population sizes seem paradoxical given the large geographical range of Neanderthals (and perhaps also of Denisovans), but they suggest that the population densities of these hominins were extremely low. Might these archaic hominins have been on their way to extinction even in the absence of any competition they may have experienced from modern humans?

The new Neanderthal genome will also provide insight into the evolution of modern humans. Prüfer *et al.* report that there are just 96 protein-coding positions at which the

Neanderthal sequence differs from that of all modern humans, with around a further 35,000 such differences at non-coding positions, some of which may affect gene regulation. This catalogue is an intriguing starting point for studying the functions of genetic differences between these groups; for example, this list is short enough to imagine creating cell lines or mouse models that contain each specific change. However, one must be mindful that many human attributes, such as bipedal gait and complex culture, probably evolved before this period of hominin diversification, and that additional important variants may lie in parts of the genome that are difficult to sequence using current methods.

After years of challenges, ancient-DNA studies are coming into their own, but they are raising as many questions as they answer. How many distinct archaic hominin groups were around in the Late Pleistocene? What were their geographical distributions? How did they help to shape the genetic make-up of modern humans? The recent sequencing of a 24,000-year-old Siberian specimen<sup>9</sup> and the recovery of mitochondrial DNA from a 400,000-year-old hominin<sup>12</sup> are examples of how each new ancient genome adds significantly to our understanding of both recent and more distant human history. We can

expect many more exciting stories in the coming years. ■

**Ewan Birney** is at the European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK. **Jonathan K.**

**Pritchard** is in the Departments of Genetics and Biology and the Howard Hughes Medical Institute, Stanford University, Stanford, California 94305-5120, USA.

e-mails: birney@ebi.ac.uk;  
pritch@stanford.edu

1. Prüfer, K. *et al.* *Nature* **505**, 43–49 (2014).
2. Finlayson, C. *et al.* *Nature* **443**, 850–853 (2006).
3. Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. & Wilson, A. C. *Science* **253**, 1503–1507 (1991).
4. Krings, M. *et al.* *Cell* **90**, 19–30 (1997).
5. Green, R. E. *et al.* *Science* **328**, 710–722 (2010).
6. Reich, D. *et al.* *Nature* **468**, 1053–1060 (2010).
7. Meyer, M. *et al.* *Science* **338**, 222–226 (2012).
8. Keller, A. *et al.* *Nature Commun.* **3**, 698 (2012).
9. Raghavan, M. *et al.* *Nature* <http://dx.doi.org/10.1038/nature12736> (2013).
10. Bachtrog, D., Thornton, K., Clark, A. & Andolfatto, P. *Evolution* **60**, 292–302 (2006).
11. Leffler, E. M. *et al.* *PLoS Biol.* **10**, e1001388 (2012).
12. Meyer, M. *et al.* *Nature* <http://dx.doi.org/10.1038/nature12788> (2013).

This article was published online on 18 December 2013.

to the IPCC's fourth (2007) and fifth (2013) assessment reports.

The researchers came up with three crucial findings. First, they observed that differences in mixing strength explained about half of the spread of climate sensitivities estimated by the models. Second, they found that changes in mixing strength depend on the mixing strength in simulations of the current climate, which was used as the initial value in the experiments. And third, they conclude that estimates of current mixing strength based on observations imply a climate sensitivity of more than 3 °C, which is in the upper half of the IPCC's range of estimates.

Another recent study<sup>4</sup> of constraints on the uncertainty of cloud responses, based on observational data, also suggested that higher climate sensitivities are more likely than lower ones. So can we declare the long-running debate about climate sensitivity to be over? Unfortunately not. Such sensitivity can also be inferred using observational data or using estimates of historical changes in surface-air temperature, heat intake by the ocean or Earth's radiative balance (the heating or cooling effects of anthropogenic greenhouse gases and aerosols). One such study, published last year, implies that climate sensitivities below 2 °C cannot be ruled out<sup>5</sup>, demonstrating that constraints on the uncertainty depend on the approaches used to determine them.

There are many factors that could explain the discrepancy. Although the uncertainty

## CLIMATE SCIENCE

# Clouds of uncertainty

**An evaluation of atmospheric convective mixing and low-level clouds in climate models suggests that Earth's climate will warm more than was thought in response to increasing levels of carbon dioxide. SEE ARTICLE P.37**

HIDEO SHIOGAMA & TOMOO OGURA

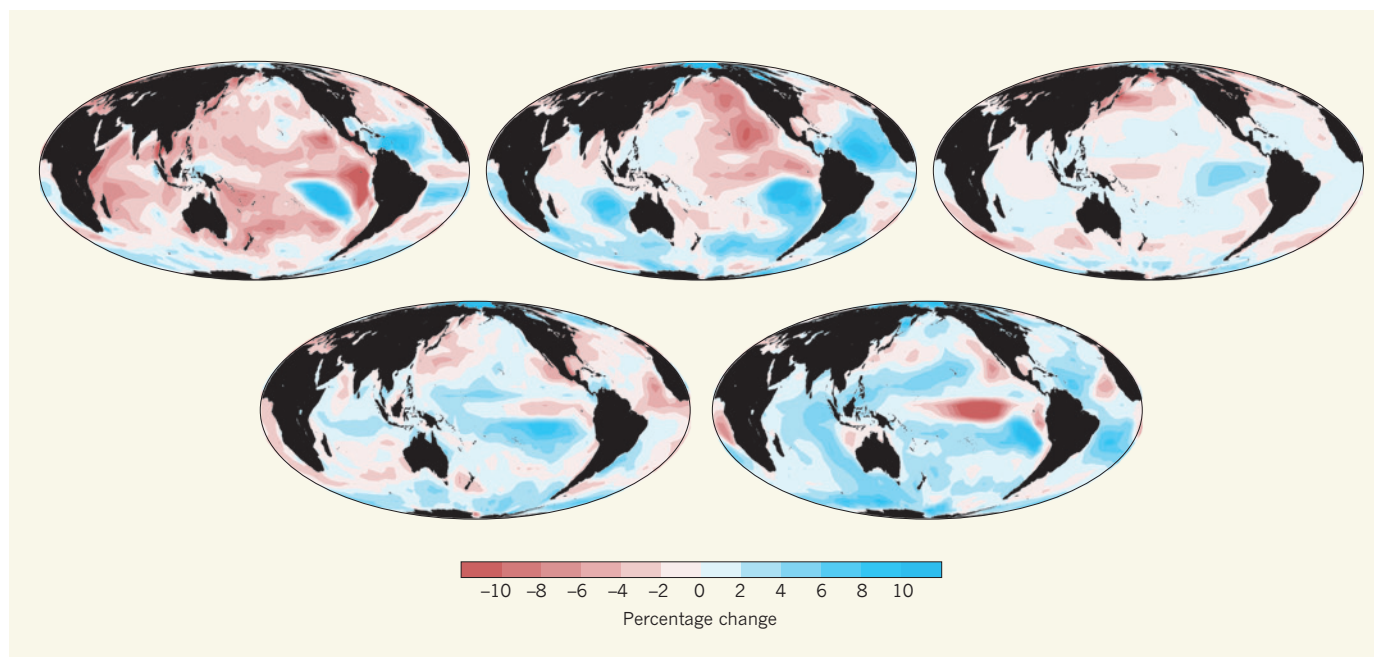
Earth is warming because of increased atmospheric concentrations of greenhouse gases, including carbon dioxide, caused by human activities. To develop policies that can help to control anthropogenic interference in climate, estimates of climate sensitivity — the mean global temperature response to a doubling of CO<sub>2</sub> levels — are required, and have been sought for decades. But despite technical advances and the considerable efforts of climate scientists, the range of climate sensitivities estimated by the Intergovernmental Panel on Climate Change (IPCC) using computer models has not narrowed since 1990, and remains at roughly 1.5–4.5 °C (ref. 1). Low-level clouds occurring below 2–3 kilometres over the tropical ocean respond in various ways to a doubling of CO<sub>2</sub> in different models<sup>2</sup> (Fig. 1), and so are key contributors to the uncertainty of climate sensitivity. On page 37 of this issue, Sherwood *et al.*<sup>3</sup>

present an observational test of atmospheric convective mixing that is relevant to low-level cloud responses, and they suggest that higher climate sensitivities are more likely than lower ones.

Low-level clouds reflect incoming sunlight from space, and so cool the climate. If the amount of this cloud declines steeply as the climate warms, then more sunlight will reach the surface, an effect that contributes to higher climate sensitivity. By contrast, increases in low-level cloud result in lower climate sensitivity.

Sherwood and colleagues propose a mechanism that controls changes in the amount of low-level cloud. They reason that, as the climate warms, stronger mixing of water vapour between the low-level cloud layer and the layer of the atmosphere above it desiccates the low-level cloud layer, reducing the amount of cloud. To assess the effect of this in climate models, the authors defined and computed measures of mixing strength for 43 models that contributed





**Figure 1 | More or less cloudy.** The maps depict estimates of changes in the coverage of low-level cloud in response to an abrupt quadrupling of  $\text{CO}_2$  levels, compared with today's levels, as determined by five of the most recent generation of climate models. Data represent mean values for the period 11–20 years after  $\text{CO}_2$  quadrupling. Sherwood *et al.*<sup>3</sup> propose processes that lead to substantial differences in the response of low-level cloud to changing  $\text{CO}_2$  levels, and which help to explain the variation in climate sensitivities calculated by the models. (Graphic generated by H.S. and T.O.)

about changes in low-level cloud over the tropical ocean contributes greatly to the uncertainty of climate sensitivity, uncertainties in other processes — such as changes in sea ice, water vapour, atmospheric temperature and cloud at other atmospheric levels and regions of the world — are also important.

Sherwood and colleagues' study represents a big advance, but questions persist. For example, around half of the spread of climate sensitivities estimated in their study remains unexplained. Furthermore, there is no guarantee that the available ensemble of climate models samples the full range of uncertainty, or that the results might not be skewed by common errors in most of the models<sup>6,7</sup>.

But although the authors' approach may not provide all the answers, the alternative approach of analysing past changes also has considerable difficulties. There are substantial uncertainties in estimates of radiative balance, and observational data on surface-air temperature and ocean heat intake suffer from limited spatial and temporal coverage, sampling biases and discontinuities associated with the use of different measurement instruments. For example, a study<sup>8</sup> last year suggests that the global warming rate in the past 15 years has been underestimated because of the lack of observations of sea surface temperatures over the Arctic region.

For now, Sherwood *et al.* have proposed and tested a convincing mechanism that explains half of the spread of models' climate sensitivities, and which suggests that future climate will be warmer than expected. The fact that their findings are variously consistent and

inconsistent with those of other studies poses further challenges for wide areas of research, including observations and reconstructions of climate systems, understanding of the processes involved, climate modelling, and analyses of climate simulations. All will be needed to solve the recondite climate-sensitivity puzzle. ■

**Hideo Shiogama and Tomoo Ogura** are at the Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba 305-8506, Japan.

#### CELL BIOLOGY

## The beginning of the end

**Studies in mice and humans suggest that cellular senescence, the cessation of cell proliferation that is known to suppress cancer and promote ageing, may have evolved to regulate embryonic development.**

JUDITH CAMPISI

Cells that experience certain types of stress, particularly stress that is potentially cancer-causing, undergo an essentially permanent arrest of proliferation termed cellular senescence<sup>1</sup>. Since its formal description in the 1960s, cellular senescence has been thought both to suppress the

*e-mails: shiogama.hideo@nies.go.jp; ogura@nies.go.jp*

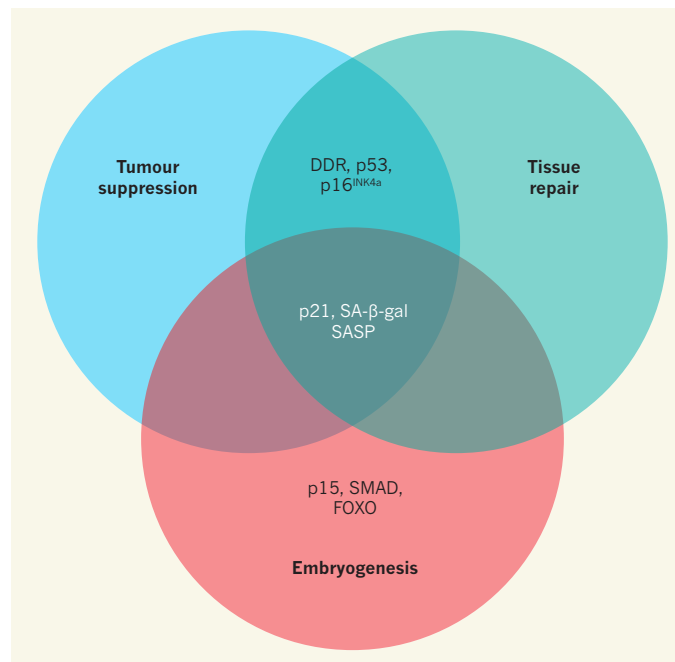
1. Jones, N. *Nature* **501**, 298–299 (2013).
2. Webb, M. J., Lambert, F. H. & Gregory, J. M. *Clim. Dyn.* **40**, 677–707 (2013).
3. Sherwood, S. C., Bony, S. & Dufresne, J.-L. *Nature* **505**, 37–42 (2014).
4. Fasullo, J. T. & Trenberth, K. E. *Science* **338**, 792–794 (2012).
5. Otto, A. *et al.* *Nature Geosci.* **6**, 415–416 (2013).
6. Knutti, R. *Clim. Change* **102**, 395–404 (2010).
7. Shiogama, H. *et al.* *Nature Commun.* **2**, 253 (2011).
8. Cowtan, K. & Way, R. G. *Q. J. R. Meteorol. Soc.* <http://dx.doi.org/10.1002/qj.2297> (2013).

development of cancer and to promote ageing. Support for these roles has come from tumour studies in mice and humans<sup>2</sup>, and from the realization that senescent cells secrete proteins that cause inflammation, a hallmark of ageing tissues<sup>3</sup>. More recently, a complex inflammatory response called the senescence-associated secretory phenotype (SASP) was shown to facilitate tissue repair and remodelling, and

to help the immune system recognize and eventually remove senescent cells<sup>4</sup>. These multiple functions of cellular senescence are not mutually exclusive, but they raise an interesting teleological question: for what purpose did senescence evolve? Findings by Muñoz-Espín *et al.*<sup>5</sup> and Storer *et al.*<sup>6</sup>, published in *Cell*, suggest a surprising answer: to fine-tune embryogenesis.

Both research groups found evidence for the presence of senescent cells in mouse and human embryos. To identify these cells, the researchers initially relied on a commonly used marker of senescence, the activity of an enzyme known as senescence-associated  $\beta$ -galactosidase (SA- $\beta$ -gal). Their combined results identified non-dividing SA- $\beta$ -gal-containing cells in the embryonic kidney, the endolymphatic sac of the inner ear, developing limbs, the closing neural tube and the apical ectodermal ridge, among other structures. Further analyses showed that non-dividing cells in these structures also expressed high levels of p21, a cell-cycle-inhibitor protein that is often expressed by senescent cells in culture and in postnatal tissues, and of a subset of SASP proteins, which are presumed to facilitate the infiltration of immune cells and eventual clearance of senescent cells (Fig. 1).

Surprisingly, however, both groups found that non-dividing cells in these embryonic structures did not express p16<sup>INK4a</sup>, a cell-cycle-inhibitor and tumour-suppressor protein that is commonly produced by senescent cells in culture and in postnatal tissues; instead, they expressed p15, another cell-cycle inhibitor that is produced by only some non-embryonic senescent cells. Similarly, the cells showed no evidence of a DNA-damage response or activation of p53, the tumour-suppressor and transcriptional-regulator protein that controls the senescence response to tissue damage or cancer-causing stress. The authors also show that senescence in the embryo depended on p21, whereas senescence in non-embryonic tissues depends primarily on p53 and p16<sup>INK4a</sup>. Moreover, p21 expression in the embryo was induced by two transcription factors, FOXO and SMAD, which are controlled by the PIK and TGF- $\beta$  signalling pathways; by contrast, induction of p21 during non-embryonic senescence is generally mediated by the DNA-damage response and p53. Thus, the senescence that occurs in embryos shares some, but not all, features of the senescence responses that suppress cancer and facilitate tissue repair (Fig. 1).



**Figure 1 | Senescence modules.** Cellular senescence is involved in tumour suppression, tissue repair and, as shown by Muñoz-Espín *et al.*<sup>5</sup> and Storer *et al.*<sup>6</sup>, embryogenesis. In all three cases, senescent cells express the protein p21 and proteins associated with the senescence-associated secretory phenotype (SASP); they also exhibit senescence-associated  $\beta$ -galactosidase (SA- $\beta$ -gal) activity. However, during tumour suppression and tissue repair, senescence depends primarily on the activity of the proteins p53 and p16<sup>INK4a</sup>, which are often induced as part of the DNA-damage response (DDR). By contrast, senescence during embryogenesis depends on p21, which is switched on by the transcription factors FOXO and SMAD. Senescent cells in the embryo also express p15.

What functions do senescent cells serve in the embryo? The authors of both papers speculate that the cells might fine-tune the development of tissue structures in the embryo, as proposed 20 years ago<sup>7</sup>. In addition to curtailing their own proliferation, senescent cells secrete factors that have potent effects on other cells<sup>4</sup>, including effects on apoptotic cell death, cell migration, immune-cell infiltration and angiogenesis (the generation of new blood vessels). It was surprising, therefore, that the researchers found only a few pre- or postnatal abnormalities in mouse embryos rendered senescence-free by deletion of the gene encoding p21. Of course, embryos are remarkably plastic and, indeed, the authors' analyses of the kinetics and structure of morphogenesis in the senescence-free embryos showed that other tissue-remodelling processes largely compensate for the lack of senescence.

The results reported by Muñoz-Espín *et al.* and Storer *et al.* are consistent with their view that cellular senescence evolved to optimize embryogenesis, and that its beneficial postnatal functions (tumour suppression and tissue repair) arose later during evolution. However, the distinct but overlapping manifestations of senescence in embryonic and postnatal tissues need not be a consequence of sequential evolution. Rather, cells might be programmed to

link arrested cell proliferation to other cellular responses, including a secretory phenotype, to meet a variety of physiological needs and respond to various forms of stress. This possibility would explain why some senescent states seem to depend primarily on p53, others on p16<sup>INK4a</sup>, yet others on p21, and so on. It might also explain why there are no markers that are unique to senescent cells<sup>4</sup>. Finally, the idea that senescence responses are assemblies of cellular characteristics might explain why the SASP proteins differ depending on the senescence inducer, cell type and tissue of origin<sup>8</sup>, including whether the senescent cells reside in postnatal tissues or the embryo<sup>5,6</sup>.

Regardless of the origin of cellular senescence, the deleterious (pro-ageing) effects of senescent cells are clearly maladaptive. The number of senescent cells increases with age in many tissues, possibly because they are incompletely eliminated by the immune system and/or produced in greater numbers in aged organisms. Consequently, aged tissues might suffer from an accumulation of non-dividing cells and the persistent presence of SASP factors that can promote chronic inflammation and alter tissue structure and function.

The findings that vertebrate embryos are replete with cells bearing characteristics of senescent cells opens up possibilities for furthering our understanding of the relationship between embryonic and adult cells, and how tissue regeneration, tumour suppression and ageing are balanced. They also raise ideas for potential therapies. Is it possible, for example, to activate embryonic senescence programs to optimize tissue repair postnatally or even in aged adults? Answers to such questions depend, of course, on future research. ■

**Judith Campisi** is at the Buck Institute for Research on Aging, Novato, California 94945, USA, and at the Lawrence Berkeley National Laboratory, Berkeley, California.  
e-mail: [jcampisi@buckinstitute.org](mailto:jcampisi@buckinstitute.org)

1. Campisi, J. & d'Adda di Fagagna, F. *Nature Rev. Mol. Cell Biol.* **8**, 729–740 (2007).
2. Collado, M. & Serrano, M. *Nature Rev. Cancer* **10**, 51–57 (2010).
3. Cevenini, E., Monti, D. & Franceschi, C. *Curr. Opin. Clin. Nutr. Metab. Care* **16**, 14–20 (2013).
4. Rodier, F. & Campisi, J. *J. Cell Biol.* **192**, 547–556 (2011).
5. Muñoz-Espín, D. *et al. Cell* **155**, 1104–1118 (2013).
6. Storer, M. *et al. Cell* **155**, 1119–1130 (2013).
7. Martin, G. M. *J. Gerontol.* **48**, B171–B172 (1993).
8. Coppé, J.-P. *et al. PLoS Biol.* **6**, e301 (2008).

This article was published online on 18 December 2013.



# Spread in model climate sensitivity traced to atmospheric convective mixing

Steven C. Sherwood<sup>1</sup>, Sandrine Bony<sup>2</sup> & Jean-Louis Dufresne<sup>2</sup>

**Equilibrium climate sensitivity refers to the ultimate change in global mean temperature in response to a change in external forcing. Despite decades of research attempting to narrow uncertainties, equilibrium climate sensitivity estimates from climate models still span roughly 1.5 to 5 degrees Celsius for a doubling of atmospheric carbon dioxide concentration, precluding accurate projections of future climate. The spread arises largely from differences in the feedback from low clouds, for reasons not yet understood. Here we show that differences in the simulated strength of convective mixing between the lower and middle tropical troposphere explain about half of the variance in climate sensitivity estimated by 43 climate models. The apparent mechanism is that such mixing dehydrates the low-cloud layer at a rate that increases as the climate warms, and this rate of increase depends on the initial mixing strength, linking the mixing to cloud feedback. The mixing inferred from observations appears to be sufficiently strong to imply a climate sensitivity of more than 3 degrees for a doubling of carbon dioxide. This is significantly higher than the currently accepted lower bound of 1.5 degrees, thereby constraining model projections towards relatively severe future warming.**

Ever since numerical global climate models (GCMs) were first developed in the early 1970s, they have exhibited a wide range of equilibrium climate sensitivities (roughly 1.5–4.5 °C warming per equivalent doubling of CO<sub>2</sub> concentration)<sup>1</sup> and consequently a broad range of future warming projections, with the uncertainty due mostly to the range of simulated net cloud feedback<sup>2,3</sup>. This feedback strength varies from roughly zero in the lowest-sensitivity models to about 1.2–1.4 W m<sup>-2</sup> K<sup>-1</sup> in the highest<sup>4</sup>. High clouds (above about 400 hPa or 8 km) contribute about 0.3–0.4 W m<sup>-2</sup> K<sup>-1</sup> to this predicted feedback because the temperatures at the tops of the clouds do not increase much in warmer climates, which enhances their greenhouse effect. Mid-level cloud changes also make a modest positive-feedback contribution in most models<sup>5</sup>.

Another positive feedback in most models comes from low cloud, occurring below about 750 hPa or 3 km, mostly over oceans in the planetary boundary layer below about 2 km. Low cloud is capable of particularly strong climate feedback because of its broad coverage and because its reflection of incoming sunlight is not offset by a commensurate contribution to the greenhouse effect<sup>6</sup>. The change in low cloud varies greatly depending on the model, causing most of the overall spread in cloud feedbacks and climate sensitivities among GCMs<sup>5,7</sup>. No compelling theory of low cloud amount has yet emerged.

A number of competing mechanisms have, however, been suggested that might account for changes in either direction. On the one hand, evaporation from the oceans increases at about 2% K<sup>-1</sup>, which—all other things being equal—may increase cloud amount<sup>8</sup>. On the other hand, detailed simulations of non-precipitating cloudy marine boundary layers show that if the layer deepens in a warmer climate, more dry air can be drawn down towards the surface, desiccating the layer and reducing cloud amount<sup>8,9</sup>.

## The lower-tropospheric mixing mechanism

We consider that a mechanism similar to this one, which has so far been considered only for a particular cloud regime, could apply more generally to shallow upward moisture transports, such as by cumulus

congestus clouds or larger-scale shallow overturning found broadly over global ocean regions. Air lifted out of the boundary layer can continue ascending, rain out most of its water vapour, and then return to a relatively low altitude—or it can exit the updraught directly at the low altitude, retaining much more of its initial vapour content. The latter process reduces the “bulk precipitation efficiency” of convection<sup>10</sup>, allowing greater transport of moisture out of the boundary layer for a given precipitation rate. Such a process can increase the relative humidity above the boundary layer<sup>11</sup> and dry the boundary layer. Unlike the global hydrological cycle and the deep precipitation-forming circulations<sup>12</sup>, however, it is not strongly constrained by atmospheric energetics<sup>11</sup>.

We present measures of this lower-tropospheric mixing and the amount of moisture it transports, and show that mixing varies substantially among GCMs and that its moisture transport increases in warmer climates at a rate that appears to scale roughly with the initial lower-tropospheric mixing.

## Mixing-induced low cloud feedback

The resulting increase in the low-level drying caused by lower-tropospheric mixing produces a mixing-induced low cloud (MILC) feedback of variable strength, which can explain why low-cloud feedback is typically positive<sup>5</sup> and why it is so inconsistent among models.

In a GCM, vertical mixing in the lower troposphere occurs in two ways (Extended Data Fig. 1). First, small-scale mixing of heat and water vapour within a single grid-column of the model is implied by convective and other parametrizations. Lower-tropospheric mixing and associated moisture transport would depend on transport by shallow cumulus clouds, but also on the downdrafts, local compensating subsidence and evaporation of falling rain that are assumed to accompany deeper cumulus. Second, large-scale mixing across isentropes occurs via explicitly resolved circulations. Whether this contributes to lower-tropospheric mixing will again depend on model parametrizations, but in this case, on their ability to sustain the relatively shallow heating that must accompany a shallow (lower-tropospheric) circulation. We measure these two mixing phenomena independently, starting with

<sup>1</sup>Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney 2052, Australia. <sup>2</sup>Laboratoire de Météorologie Dynamique and Institut Pierre Simon Laplace (LMD/IPSL), CNRS, Université Pierre et Marie Curie, Paris 75252, France.

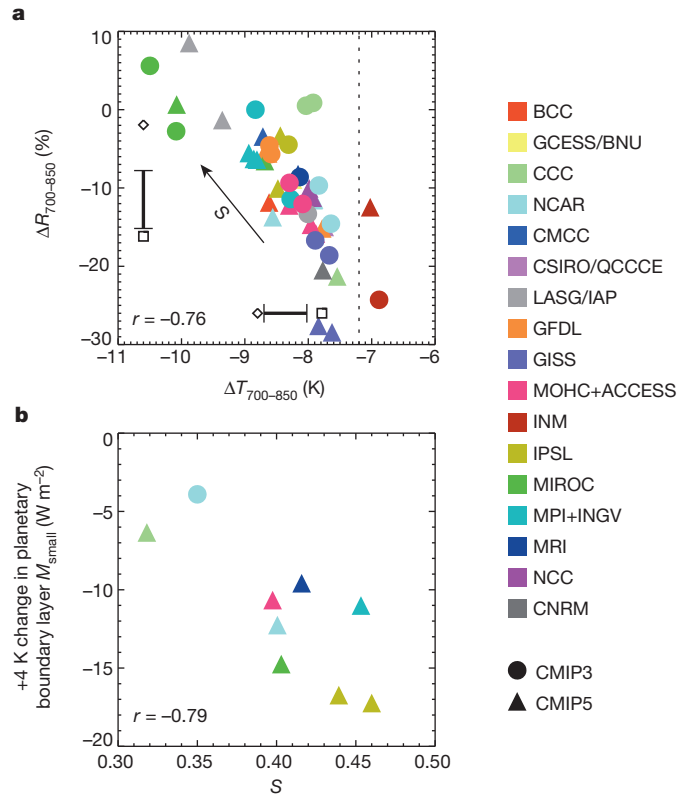
the small-scale part, and show that both phenomena progressively dry the boundary layer as climate warms.

### The small-scale component of mixing

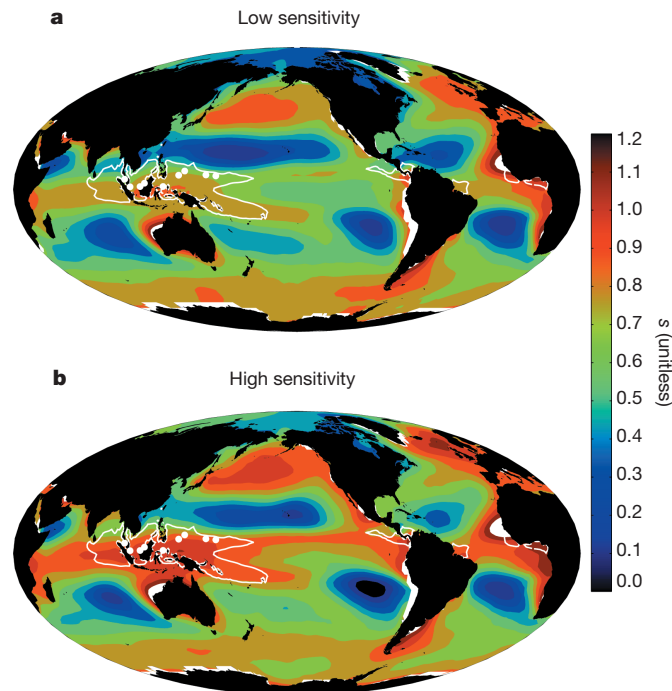
Lower-tropospheric mixing parametrized within a GCM grid cell cannot be directly diagnosed from model output (although it contributes to the convective terms in the water vapour budget; see below). We assert, however, that an atmosphere's propensity to generate such mixing can be gauged by observing the thermal structure just above the boundary layer in ascending, raining regions. As discussed above, air there is either transported directly from the boundary layer with minimal precipitation via lower-tropospheric mixing, or indirectly by ascending in deeper, raining clouds and then descending. The air would arrive cool and humid in the former case, but warmer and drier in the latter case owing to the extra condensation, allowing us to evaluate which pathway dominates by observing mean-state air properties.

To do this we use an index  $S$ , proportional to the differences  $\Delta T_{700-850}$  and  $\Delta R_{700-850}$  of temperature and relative humidity between 700 hPa and 850 hPa ( $S$  taken as a linear combination; see Methods Summary) averaged within a broad ascending region which roughly coincides with the region of highest Indo-Pacific ocean temperatures (the Indo-Pacific Warm Pool; Fig. 1). Of the full set of 48 models used in this study, those with a less negative  $\Delta T_{700-850}$  in this region consistently show a more negative  $\Delta R_{700-850}$  there (Fig. 2a), and the variations in each quantity are quite large. We interpret this as strong evidence that both quantities are dominated by variations, evidently large, in the amount of lower-tropospheric mixing in the ascent region, with higher  $S$  indicating stronger mixing.

Small-scale lower-tropospheric mixing of moisture is part of the overall source of the water vapour that is associated with the parametrized convection,  $M_{\text{small}}$ . This quantity is available from nine of the models (see Methods Summary). It always exhibits strong drying



**Figure 2 | Basis for the index  $S$  of small-scale lower-tropospheric mixing and its relationship to the warming response.** **a**,  $\Delta T_{700-850}$  versus  $\Delta R_{700-850}$ , each averaged over a tropical region of mean ascent (see Fig. 1), from all 48 coupled models. For reference, a saturated-adiabatic value of  $\Delta T$  is shown by dotted line at  $-7.2$  K, and a dry-adiabatic value (not shown) would be about  $-16$  K. Error bars are  $2\sigma$  ranges. **b**, Change in small-scale moisture source  $M_{\text{small}}$  below 850 hPa in the tropics upon  $+4$  K ocean warming, versus  $S$  computed from the control run, in eight atmosphere models and one CMIP3 model. Symbol colour indicates modelling centre or centre where atmosphere model was originally developed and symbol shape indicates model generation.



**Figure 1 | Multimodel-mean local stratification parameter  $s$ .** The index  $S$  is the mean of  $s$  within the regions outlined in white. Multimodel averages of  $s$  are shown separately for low-sensitivity ( $\text{ECS} < 3.0$  °C) (**a**) and high-sensitivity ( $\text{ECS} > 3.5$  °C) (**b**) models, among coupled models with known ECS. The white dots inside the  $S$ -averaging region show the locations of radiosonde stations used to help estimate  $S$  observationally. A few coastal regions that are off-scale appear white.

near the surface. Above about 850 hPa, it can either dry the atmosphere on average or moisten it depending on the model (Extended Data Fig. 2), reflecting the competition between drying from condensation and moistening from lower-tropospheric mixing and from evaporating precipitation falling from higher altitudes.

Although  $M_{\text{small}}$  does not reflect lower-tropospheric mixing alone, we can test whether lower-tropospheric mixing (as diagnosed from  $S$ ) affects how  $M_{\text{small}}$  responds as climate warms. The available data confirm that, given a  $+4$  K warming, convective drying of the planetary boundary layer increases by  $4\text{--}17$   $\text{W m}^{-2}$  ( $6\text{--}30\%$ ), compared to a typical increase of  $8\%$  in global or tropical surface evaporation. The drying increase is highly correlated ( $r = -0.79$ ) with  $S$  (Fig. 2b). Thus, convective dehydration of the planetary boundary layer outstrips the increase in surface evaporation with warming, in all models except those with the lowest  $S$ . Higher-sensitivity models also have higher  $S$  (Fig. 1), suggesting that this process drives a positive feedback on climate.

### The large-scale component of mixing

We next turn to the large-scale lower-tropospheric mixing, which we associate with shallow ascent or flows of air upward through the top of the boundary layer that diverge horizontally before reaching the upper troposphere. Although air ascending on large scales over warm tropical oceans typically passes through nearly the whole troposphere, over cooler oceans its ascent often wanes with altitude, showing that this type of mixing indeed occurs in the Earth's atmosphere (Fig. 3). The associated mid-level outflows are well documented for the central



and eastern Pacific and Atlantic Intertropical Convergence Zone and some monsoon circulations<sup>13,14</sup>. Although these are indeed the regions where shallow ascent is steadiest, and hence clearest in monthly-mean data (Fig. 3), in daily reanalysis data, shallow ascent is equally strong outside the tropics owing largely to contributions from extratropical storms. We also note that although we focus here on regions of ascending air, that is because the ascending branches are where the circulations are easiest to measure; they must, however, descend elsewhere, exerting a net transport of water vapour that is upward and towards the convective regions.

Figure 3 compares the observations with two example models. Neither model shows as much shallow ascent (red colour) as the observation-based estimates, but the Institut Pierre Simon Laplace (IPSL)-CM5A model comes closer. Although convective treatment in the newer IPSL-CM5B model is more detailed and produces better results in important respects<sup>15</sup>, here it is seen to produce strong deep ascent of air (white spots) where it is weaker and shallower in observations (red zones), showing that improvement in some aspects of a simulation does not automatically improve others.

We quantify the large-scale lower-tropospheric mixing more thoroughly by calculating the ratio  $D$  of shallow to deep overturning (see Methods Summary) in a broad region encompassing most of the persistent shallow ascent (see Fig. 3). This index  $D$  varies by a factor of four across 43 GCMs (see below). Interestingly, however,  $D$  and  $S$  are uncorrelated ( $r = 0.01$ ), confirming that the two scales of mixing are controlled by different aspects of model design.

The effective source of moisture  $M_{LT,large}$  due to this shallow overturning (that is, large-scale, lower-tropospheric convection) and its change upon climate warming, can be directly calculated from model wind and humidity fields. We approximate  $M_{LT,large}$  using monthly-mean data from the ten available atmospheric models (see Methods Summary).  $M_{LT,large}$  isolates only shallow mixing, whereas  $M_{small}$

includes the effects of all parameterized convection; yet despite this, the profiles  $M_{LT,large}$  (Fig. 4) resemble those of  $M_{small}$ , with strong drying in the boundary layer and weak moistening above. Not unexpectedly, these effects are greater in the high- $D$  models than in the low- $D$  ones.

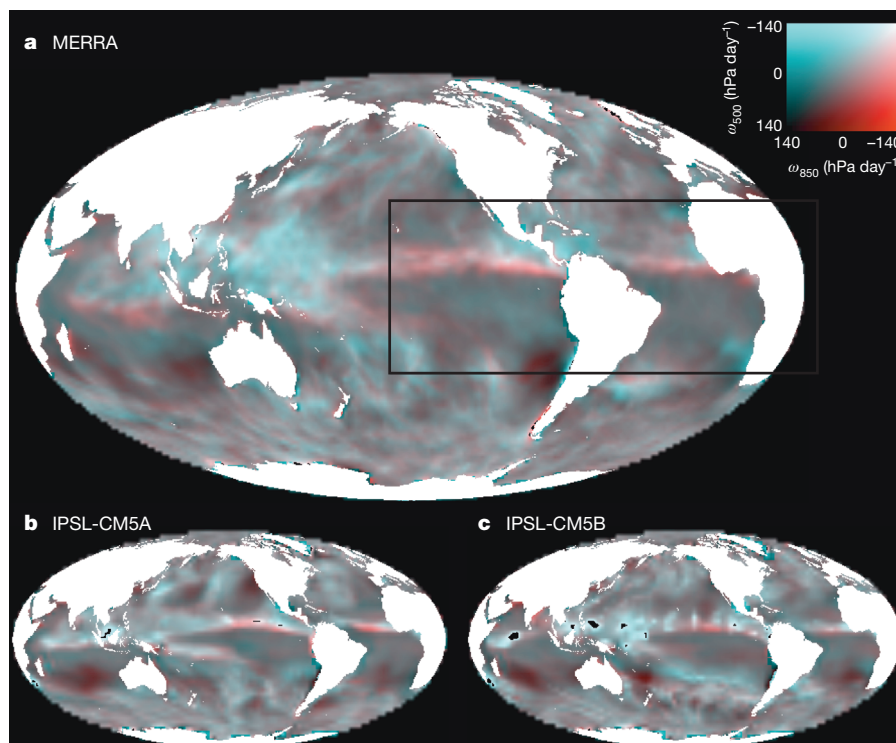
Crucially, the low-level drying also increases faster upon +4 K warming in the high- $D$  models (by about 30%, or  $1.5 \text{ W m}^{-2} \text{ K}^{-1}$  when expressed as a latent heat flux) than in the low- $D$  models (25%, or  $0.9 \text{ W m}^{-2} \text{ K}^{-1}$ ). Thus, the response of  $M_{LT,large}$  grows with  $D$  as  $M_{small}$  grew with  $S$ ; the relationship for  $D$  is not as strong ( $r = 0.46$  for land + ocean,  $r = 0.25$  for ocean only), partly because the spread of  $D$  happens to be somewhat narrow among the available atmosphere models, but is still significant at 95% confidence.

### Climate sensitivity

We now apply the indices  $S$  and  $D$  to the 43 GCMs for which an equilibrium climate sensitivity (ECS) is available. Each index independently explains about 25% of the variance in ECS (Fig. 5a, b).

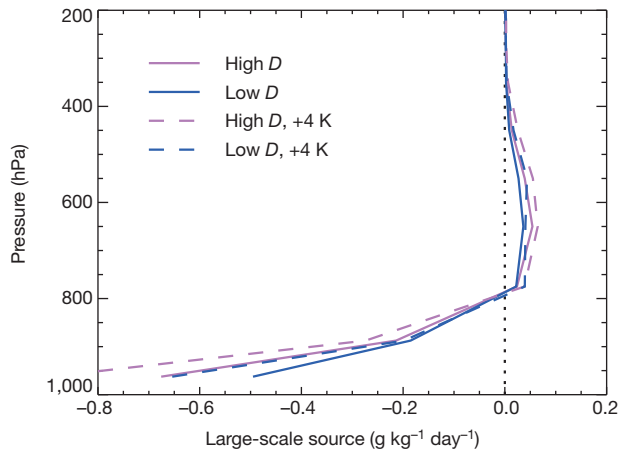
Because the ranges of  $D$  and  $S$  are similar (each 0.3–0.4), as are (approximately) those of their drying responses upon warming (see below), we form an overall lower-tropospheric mixing index (the LTMI) by simply adding the two:  $\text{LTMI} = S + D$ . This LTMI explains about 50% of the variance in total system feedback ( $r = 0.70$ ) and ECS ( $r = 0.68$ ) (Fig. 5c). Thus, although our measure of lower-tropospheric mixing does not explain all of the variations among GCMs, it does explain a significant portion of the model spread.

This explanatory power derives primarily from low cloud feedbacks. The correlation between LTMI and the +4 K change in short-wave cloud radiative effect in the atmosphere models, which spans a range of  $1.8 \text{ W m}^{-2} \text{ K}^{-1}$  in the tropics, is 0.65 in the tropics and 0.57 in subsidence regions (equivalent values estimated from a subset of the coupled models providing the needed output are 0.25 and 0.47



**Figure 3 | The structure of monthly-mean tropospheric ascent reveals large-scale lower-tropospheric mixing in observations and models.** Upward pressure velocity  $\omega$  in one month (September) from the MERRA reanalysis (a), the IPSL-CM5A model (b) and the IPSL-CM5B model (c) with values at 850 hPa shown in red and those at 500 hPa shown in green plus blue. Bright red implies ascent that is weighted toward the lower troposphere with

mid-tropospheric divergence (see colour scale), white implies deep ascent, and dark colours imply descent. In a, black lines outline the region in which the index  $D$  of large-scale lower-tropospheric mixing is computed. The Pacific and Atlantic Intertropical Convergence Zone regions are consistently red in the reanalyses and models, whereas isolated red patches in other areas tend to vary with time.

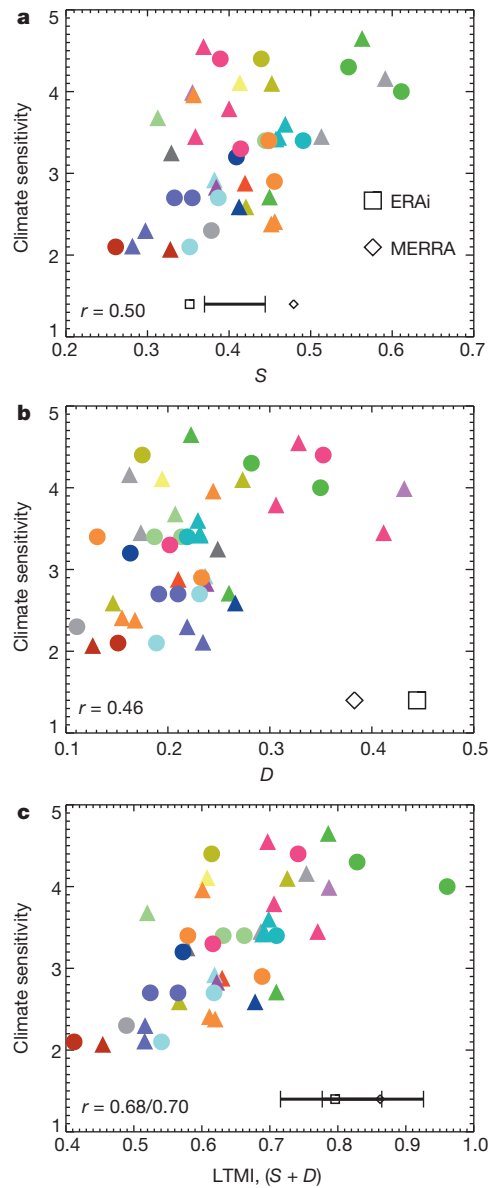


**Figure 4 | Estimated water vapour source  $M_{LT,large}$  due to large-scale lower-tropospheric mixing and its response to warming.** See Methods for calculation details. Data are from ten atmosphere models, averaged from 30° S to 30° N over oceans, with the average of the four models having the largest  $D$  shown in magenta and the average of the four models with the smallest  $D$  shown in blue. Dashes show results in +4 K climate. Changes at +4 K are nearly identical whether or not land areas are included.

respectively). These correlations suggest that the predictive skill of LTMI arises from both subsidence and other regions; further work is needed to better assess this. Cloud amount reduces more in high-LTMI models both at low and mid-levels (Extended Data Fig. 3), although the greater net radiative impact of low cloud makes its effect dominant<sup>16</sup>. Previously reported water vapour and lapse-rate feedbacks<sup>17</sup> are, in contrast, not correlated with the LTMI.

Is the imputed lower-tropospheric mixing impact on low clouds strong enough to explain the approximately  $1.5 \text{ W m}^{-2} \text{ K}^{-1}$  spread of cloud feedbacks seen in GCMs?<sup>4</sup> One recent study<sup>18</sup> imposed increased surface latent heat fluxes in a large region typified by shallow clouds, finding an increase in cloud-related net cooling of about  $1 \text{ W m}^{-2}$  for a  $2\text{--}3 \text{ W m}^{-2}$  increase in the surface flux, other things held fixed. An even larger sensitivity, nearly 1:1, has been reported in a different model for advective changes in moisture input<sup>19</sup>. If a similar but opposite cloud response occurred for moisture removal by lower-tropospheric mixing, then to explain the feedback spread, the boundary-layer drying responses would need to span a range across models of about  $3 \text{ W m}^{-2}$  per K of surface warming. This roughly matches the contribution to the spread from  $M_{small}$  alone (Fig. 2b). The additional drying response from  $M_{LT,large}$  was about  $0.6 \text{ W m}^{-2} \text{ K}^{-1}$  greater in the high- $D$  models (mean  $D$  of 0.34) than in the low- $D$  ones (mean 0.24), which, if rescaled by the full spread of  $D$  in the full GCM ensemble, implies a further source of spread in drying response of about  $2 \text{ W m}^{-2} \text{ K}^{-1}$ . We conclude that, even if not all low clouds are as sensitive as the ones examined in the cited studies, the lower-tropospheric mixing response is strong enough to account for the cloud feedback spread and its typically positive sign<sup>5</sup>.

Why does moisture transport increase so strongly with warming? The magnitude of these increases, typically 5%–7% per K of surface warming, is roughly what would be expected if the circulations remained similar against a Clausius–Clapeyron increase in moisture gradients<sup>20</sup>, as indeed it does, at least for the large-scale part<sup>21</sup> (Extended Data Fig. 4). Further study is needed to understand why this is so, and to examine in greater detail how clouds respond to changing moisture transports; changes in low cloud amount may for example help the atmosphere restore imbalances in boundary layer moist enthalpy such as those caused by lower-tropospheric mixing<sup>19</sup>. Because LTMI ignores any information on clouds, it is likely that additional measures of cloud characteristics<sup>22</sup> could explain some of the variations in low-cloud feedback not yet explained here.



**Figure 5 | Relation of lower-tropospheric mixing indices to ECS.** ECS versus  $S$  (a),  $D$  (b) and  $LTMI = S + D$  (c) from the 43 coupled models with known ECS. Linear correlation coefficients  $r$  are given in each panel ( $r = 0.70$  in c is the correlation to the total system feedback). Error bars shown near panel axes indicate  $2\sigma$  ranges of the direct radiosonde estimate (a) and the  $S$  value from radiosondes added to the  $D$  value from each of the two reanalyses (c). ERAi and MERRA are the two monthly reanalysis products.

We end by considering observational estimates of  $S$  and  $D$  (see Fig. 5). These show an  $S$  near the middle of the GCM range, but a  $D$  close to the top end, as hinted already by Fig. 3.  $D$  may not be well constrained because  $\omega$  must be inferred from observational reanalyses, although available horizontal wind observations support the existence of strong mid-level outflows<sup>13</sup>, and the result is consistent across both reanalyses examined. The reanalysis estimates of  $S$  are less consistent but this quantity can be fairly well constrained by radiosonde observations.

Taking the available observations at face value implies a most likely climate sensitivity of about  $4^\circ \text{C}$ , with a lower limit of about  $3^\circ \text{C}$ . Indeed, all 15 of the GCMs with ECS below  $3.0^\circ \text{C}$  have an LTMI below the bottom of the observational range. Further work may be needed to better constrain these indices, and to test whether their relationship to ECS is robust to design factors common to all models. For example, this should be tested in global cloud-resolving models.



The possibility can never be ruled out that feedbacks could exist in nature that are missing from all models, which would change the climate sensitivity from that suggested by our result. Nonetheless, on the basis of the available data, the new understanding presented here pushes the likely long-term global warming towards the upper end of model ranges.

## Discussion

Although a few previous studies have already noted that higher-sensitivity models simulate certain cloud-relevant phenomena better<sup>23–25</sup>, ours is the first to demonstrate a causal physical mechanism, or to show consistent predictive skill across so many models, or to point to processes connecting low-cloud regions to the deep tropics. The MILC mechanism is surprisingly straightforward. Lower-tropospheric mixing dries the boundary layer, and the drying rate increases by 5–7% K<sup>−1</sup> in warmer climates owing to stronger vertical water vapour gradients. The moisture source from surface evaporation increases at only about 2% K<sup>−1</sup>. Thus as climate warms, any drying by lower-tropospheric mixing becomes larger relative to the rest of the hydrological cycle, tending to dry the boundary layer. How important this is depends on how important the diagnosed lower-tropospheric mixing was in the base state of the atmosphere. Lower-tropospheric mixing is unrealistically weak in models that have low climate sensitivity.

Climate-sensitivity-related differences in lower-tropospheric mixing, both at small (Fig. 1) and large scales (Fig. 3), are most detectable in regions of tropical deep or mixed-level convection and mean upward motion. This does not mean, however, that the greater low-level drying in a warmer climate or the spread of drying among models will be limited to these regions.

Large-scale lower-tropospheric mixing carries water vapour not only upward but also horizontally away from subsidence regions; because both directions of transport intensify in a warmer atmosphere<sup>20</sup>, subsidence regions should bear the brunt of the overall boundary-layer drying. Moreover, shallow ascent is equally strong (though more transient) in mid-latitude storm tracks and in the tropics, suggesting that MILC feedback may be just as important outside the tropics as within them.

As for small-scale lower-tropospheric mixing, even though there are reasons to measure it in ascending regions (see Methods), its impact upon warming is much more widespread and differs significantly among models in subsiding regions (Extended Data Fig. 5). We hypothesise that this is because models with more small-scale lower-tropospheric mixing in ascending regions also have more in descending regions, although we cannot confirm this directly. Overall, the behaviour is consistent with published results showing that subsiding regions contribute strongly to the spread of cloud feedbacks in models, with storm tracks and tropical convective regions also playing a part<sup>16,26,27</sup>.

Lower-tropospheric mixing behaviour appears to result from a competition between shallow and deep convection in situations where either could occur. Such situations persist in many tropical regions, notably the Intertropical Convergence Zone. Understanding and properly representing this competition in climate models is undoubtedly necessary for more accurate future climate projections.

Although tested here on models used over the past decade or so, we presume that this mechanism has been a leading source of spread in sensitivity since the dawn of climate modelling. Finally to identify an atmospheric process that drives variations in climate sensitivity offers an unprecedented opportunity to focus research and model development in ways that should lead to more reliable climate change assessments.

## METHODS SUMMARY

Data for computing  $S$  and  $D$  come from control runs of 48 models: 18 from the Coupled Model Intercomparison Project version 3 (CMIP3)<sup>28</sup> and 30 from CMIP5 (ref. 29) (see Extended Data Tables 1 and 2). ECS was reported for all but one CMIP3 model by the Intergovernmental Panel on Climate Change<sup>28</sup>. For CMIP5 we employ effective climate sensitivities calculated from abrupt 4 × CO<sub>2</sub> experiments, available for 26 models, following a standard regression procedure<sup>30,31</sup>.

Data for  $M_{\text{small}}$  and  $M_{\text{LT,large}}$  come from ten CMIP5 atmosphere models providing ‘amip’ (specified ocean surface temperature) control and +4 K ocean warming runs. Eight of these models provided  $M_{\text{small}}$ ; we also included data from the Parallel Climate Model (CMIP3).

Observational estimates come from radiosondes and two monthly reanalysis products (ERAi and MERRA). Reanalyses are produced from a model constrained to the fullest extent possible by a variety of observations<sup>32,33</sup>.

We calculate  $S$  within a region where convective effects are a leading term in thermodynamic budgets, defined by the upper quartile of the annual-mean mid-tropospheric ascent rate where it is upward,  $-\omega_{500}$  ( $\omega$  the pressure velocity). We define  $S \equiv (\Delta R_{700-850}/100\% - \Delta T_{700-850}/9 \text{ K})/2$ , which normalizes  $\Delta R_{700-850}$  to 100% humidity,  $\Delta T_{700-850}$  to the approximately 9-K range between dry and saturated adiabatic values, and averages these two pieces of information with equal weight to reduce noise from other factors.

To calculate  $M_{\text{LT,large}}$  we compute  $\omega_1$  (the average of  $\omega$  at 850 hPa and 700 hPa) and  $\omega_2$  (the average of  $\omega$  at 600 hPa, 500 hPa and 400 hPa).  $\Delta = \omega_2 - \omega_1$  measures the local horizontal outflow in the lower troposphere above the boundary layer. Moisture is transported upward and outward wherever  $\Delta > 0$  and  $\omega_1 < 0$ . We restrict measurement to tropical ocean regions from 160° W to 30° E (see Fig. 3). The moisture supplied to the environment is estimated as  $M_{\text{LT,large}} = -\langle q d\omega / dp H(\Delta) H(-\omega_1) \rangle$ , where  $p$  is the pressure,  $q$  is the specific humidity,  $\langle \dots \rangle$  indicates the mean over the restricted region, and  $H$  is the step function. Finally,  $D \equiv \langle \Delta H(\Delta) H(-\omega_1) \rangle / \langle -\omega_2 H(-\omega_2) \rangle$ .

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 May; accepted 5 November 2013.

- Hegerl, G. C. & Zwiers, F. W. *Working Group I: The Physical Science Basis of Climate Change* Ch. 9, 663–745 (Cambridge Univ. Press, 2007).
- Cess, R. D. *et al.* Intercomparison and interpretation of climate feedback processes in 19 atmospheric general circulation models. *J. Geophys. Res.* **95**, 16601–16615 (1990).
- Dufresne, J. L. & Bony, S. An assessment of the primary sources of spread of global warming estimates from coupled atmosphere-ocean models. *J. Clim.* **21**, 5135–5144 (2008).
- Soden, B. J. *et al.* Quantifying climate feedbacks using radiative kernels. *J. Clim.* **21**, 3504–3520 (2008).
- Zelinka, M., Klein, S. & Hartmann, D. Computing and partitioning cloud feedbacks using cloud property histograms. Part I: cloud radiative kernels. *J. Clim.* **25**, 3715–3735 (2012).
- Hartmann, D. L. & Short, D. A. On the use of earth radiation budget statistics for studies of clouds and climate. *J. Atmos. Sci.* **37**, 1233–1250 (1980).
- Bony, S. & Dufresne, J. L. Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophys. Res. Lett.* **32**, L20806 (2005).
- Rieck, M., Nuijens, L. & Stevens, B. Marine boundary-layer cloud feedbacks in a constant relative humidity atmosphere. *J. Atmos. Sci.* **69**, 2538–2550 (2012).
- Stevens, B. On the growth of layers of nonprecipitating cumulus convection. *J. Atmos. Sci.* **64**, 2916–2931 (2007).
- Yano, J. & Emanuel, K. An improved model of the equatorial troposphere and its coupling with the stratosphere. *J. Atmos. Sci.* **48**, 377–389 (1991).
- Sherwood, S. C. & Meyer, C. L. The general circulation and robust relative humidity. *J. Clim.* **19**, 6278–6290 (2006).
- Allen, M. R. & Ingram, W. J. Constraints on future changes in climate and the hydrologic cycle. *Nature* **419**, 224–232 (2002).
- Zhang, C. D., Nolan, D. S., Thorncroft, C. D. & Nguyen, H. Shallow meridional circulations in the tropical atmosphere. *J. Clim.* **21**, 3453–3470 (2008).
- Nolan, D. S., Powell, S. W., Zhang, C. D. & Mapes, B. E. Idealized simulations of the Intertropical Convergence Zone and its multilevel flows. *J. Atmos. Sci.* **67**, 4028–4053 (2010).
- Rio, C., Hourdin, F., Grandpeix, J. Y. & Lafore, J. P. Shifting the diurnal cycle of parameterized deep convection over land. *Geophys. Res. Lett.* **36**, L07809 (2009).
- Soden, B. J. & Vecchi, G. A. The vertical distribution of cloud feedback in coupled ocean-atmosphere models. *Geophys. Res. Lett.* **38**, L12704 (2011).
- Soden, B. J. & Held, I. M. An assessment of climate feedbacks in coupled ocean-atmosphere models. *J. Clim.* **19**, 3354–3360 (2006).
- Webb, M. & Lock, A. Coupling between subtropical cloud feedback and the local hydrological cycle in a climate model. *Clim. Dyn.* **41**, 1923–1939 (2013).
- Brient, F. & Bony, S. Interpretation of the positive low-cloud feedback predicted by a climate model under global warming. *Clim. Dyn.* **40**, 2415–2431 (2013).
- Held, I. M. & Soden, B. J. Robust responses of the hydrological cycle to global warming. *J. Clim.* **19**, 5686–5699 (2006).
- Zahn, M. & Allan, R. P. Climate warming-related strengthening of the tropical hydrological cycle. *J. Clim.* **26**, 562–574 (2013).
- Klein, S. A. *et al.* Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator. *J. Geophys. Res.* **118**, 1329–1342 (2013).

23. Volodin, E. M. Relation between temperature sensitivity to doubled carbon dioxide and the distribution of clouds in current climate models. *Izv. Atmos. Ocean. Phys.* **44**, 288–299 (2008).
24. Clement, A. C., Burgman, R. & Norris, J. R. Observational and model evidence for positive low-level cloud feedback. *Science* **325**, 460–464 (2009).
25. Fasullo, J. T. & Trenberth, K. E. A less cloudy future: the role of subtropical subsidence in climate sensitivity. *Science* **338**, 792–794 (2012).
26. Webb, M. J., Lambert, F. H. & Gregory, J. M. Origins of differences in climate sensitivity, forcing and feedback in climate models. *Clim. Dyn.* **40**, 677–707 (2013).
27. Vial, J., Dufresne, J.-L. & Bony, S. On the interpretation of inter-model spread in CMIP5 climate sensitivity estimates. *Clim. Dyn.* **41**, 3339–3362 (2013).
28. Randall, D. A. & Wood, R. A. *Working Group I: The Physical Science Basis of Climate Change* Ch. 8, 589–662 (Cambridge University Press, 2007).
29. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
30. Gregory, J. M. *et al.* A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.* **31**, L03205 (2004).
31. Forster, P. M. *et al.* Evaluating adjusted forcing and model spread for historical and future scenarios in the CMIP5 generation of climate models. *J. Geophys. Res.* **118**, 1139–1150 (2013).
32. Uppala, S. M. *et al.* The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**, 2961–3012 (2005).
33. Rienecker, M. M. *et al.* MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Clim.* **24**, 3624–3648 (2011).

**Acknowledgements** This work was supported by the FP7-ENV-2009-1 European project EUCLIPSE (number 244067). We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output, especially the participants contributing additional CFMIP2 experiments and diagnostics crucial to our study. The US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support for CMIP and led the development of software infrastructure in partnership with the Global Organisation for Earth System Science Portals. We also thank the National Center for Atmospheric Research and the Earth System Grid Federation for providing access to PCM output, the Australian National Computational Infrastructure, and the IPSL Prodiguer-Ciclad facility for providing a convenient archive of CMIP data. Finally, we thank B. Stevens, C. Bretherton and G. Schmidt for comments on early versions of the manuscript.

**Author Contributions** S.C.S. led the study and the writing of the paper, and did the calculations of LTMI and related diagnostics. S.B. computed cloud radiative effect and assisted in interpreting results and writing the paper. J.-L.D. computed ECS and assisted in interpreting results and writing the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.C.S. ([s.sherwood@unsw.edu.au](mailto:s.sherwood@unsw.edu.au)).



## METHODS

Data for computing  $S$  and  $D$  come from 48 models: 18 from the CMIP3 (Coupled Model Intercomparison Project version 3)<sup>28</sup>, the first two years of each “picntrl” run, and 30 models from the CMIP5 (ref. 29), the first two years of each “1pctCO2” run. Two years of data is sufficient to specify  $S$  and  $D$  to within 0.02 or better of their long-term values. CMIP3 data were obtained from the Australian National Computational Infrastructure node, and CMIP5 data including the ‘amip’ and ‘amip+4K’ runs were obtained on 14 September 2012 and 22 October 2012 from the IPSL Cicalad repository. ECS values for CMIP3 were reported for all but one model by the Intergovernmental Panel on Climate Change<sup>28</sup>. For CMIP5 we employ effective climate sensitivities calculated from abrupt  $4 \times \text{CO}_2$  experiments, available for 26 of the 30 CMIP5 models, following a standard regression procedure<sup>30,31</sup>.

Data for  $M_{\text{small}}$  and  $M_{\text{LT, large}}$  come from ten CMIP5 atmosphere models providing ‘amip’ (specified ocean surface temperature) control and +4 K ocean warming experiments. A key advantage of this experiment setup is that inter-annual ocean variability is the same in the control and warming runs, and changes in the sea surface temperature pattern—which could complicate interpretation, especially for circulation changes—are avoided. Data are from 1989–98, except for IPSL-CM5A, in which some of these years were corrupted and alternative years were used. Results from individual years were similar to those for the ten-year averages. Eight of these models provided  $M_{\text{small}}$ ; we also included data from the PCM CMIP3 1%-per-year-to-quadrupling experiment, with changes rescaled to the +4 K equivalent (actual change 3.3 K). PCM  $M_{\text{small}}$  data come from ten years near the beginning and ten years near the end of the 1%-per-year-to-quadrupling experiment, obtained from the National Center for Atmospheric Research node of the Earth System Grid.

The shortwave cloud radiative effect is obtained by differencing the all-sky and clear-sky top-of-atmosphere shortwave fluxes for each model run. To calculate cloud feedback we first composite the sensitivity of the shortwave cloud radiative effect to sea surface temperature in dynamical regimes defined by vertical-mean vertical velocity, and then we compute the sum (weighted by the probability distribution function of  $\omega$ ) over regimes (or only subsidence regimes defined by  $\omega > 0$ )<sup>7</sup>. For coupled models, the warming-induced change is obtained from abrupt  $\text{CO}_2$ -quadrupling experiments, after removing the instantaneous change associated with rapid adjustment to higher  $\text{CO}_2$  estimated from the first 12 months after quadrupling. Only one realization is used per model. For atmosphere-only models it is simply the difference between the +4 K and the control simulations.

Observational estimates come from radiosondes and from two monthly reanalysis products (ERAi and MERRA), years 2009–10. The reanalyses are produced from a model constrained to the full extent possible by a variety of observations<sup>32,33</sup>. MERRA reanalysis data from 1 September 2009 were used to compare  $D$  inside and outside the tropics, but monthly data were used otherwise. Radiosonde data were obtained from the Integrated Global Radiosonde Archive and subjected to simple quality-control checks for outliers. The ten stations sited in the relevant region and meeting the criteria described by a previous study<sup>34</sup> were used, and the mean taken over the 2 years. The radiosonde network sampling bias, as determined from station-sampled reanalysis output, was relatively small compared to the overall reanalysis biases.

We calculate  $S$  in ascending regions, where convective effects are a leading term in thermodynamic budgets; in subsidence regions humidity is sensitive to irrelevant non-local factors and even to numerical resolution<sup>35</sup>, perhaps explaining why it is less informative for our purposes. The calculation region is defined by the upper quartile of the annual-mean mid-tropospheric ascent rate in ascending regions,  $-\omega_{500}$  (where  $\omega$  is the pressure velocity). We define  $S \equiv (\Delta R_{700-850}/100\% - \Delta T_{700-850}/9 \text{ K})/2$ , which normalizes  $\Delta R_{700-850}$  to 100% humidity and  $\Delta T_{700-850}$  to the approximately 9 K range between the dry and saturated adiabatic values, and then averages these two pieces of information with equal weight. Such averaging should reduce the noise from other factors that influence one quantity or the other. Varying the weighting of the two terms does not strongly affect results.

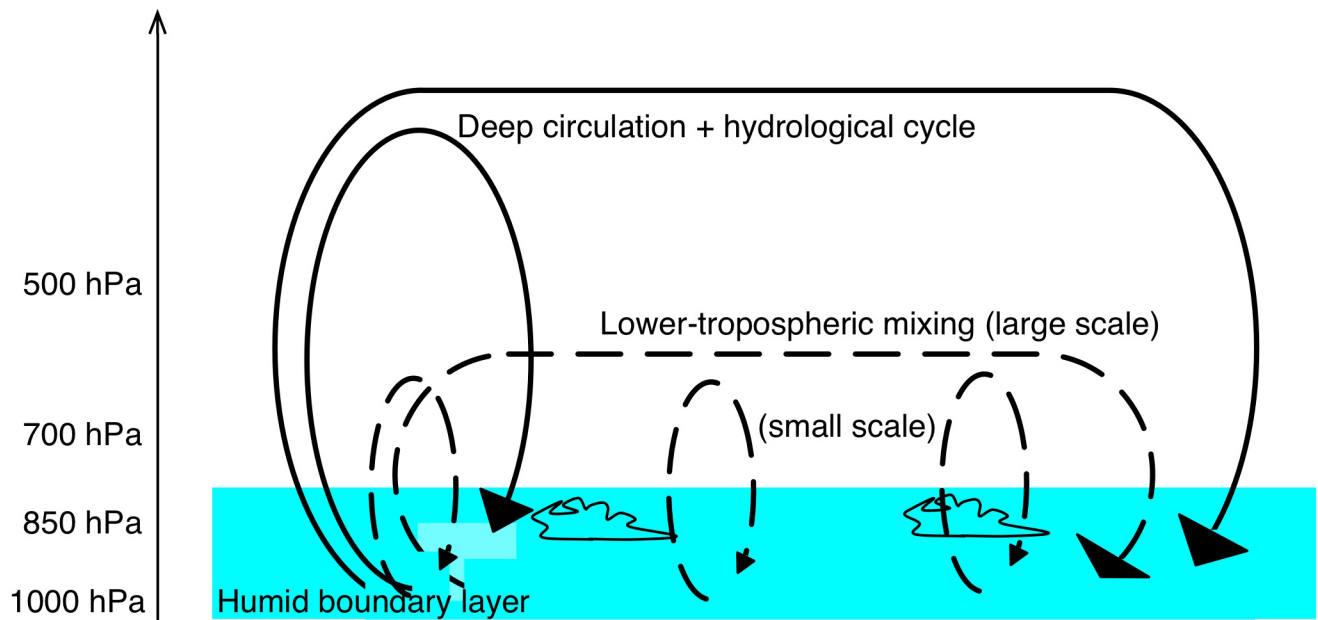
To calculate  $M_{\text{LT, large}}$ , we first compute  $\omega_1$  (the average  $\omega$  at 850 hPa and 700 hPa) and  $\omega_2$  (the average  $\omega$  at 600 hPa, 500 hPa and 400 hPa). The difference  $\Delta = \omega_2 - \omega_1$  then measures the local horizontal outflow in the lower troposphere above the boundary layer. Moisture is transported upward and outward at this level wherever  $\Delta > 0$  and  $\omega_1 < 0$ . We restrict measurement to tropical ocean regions from 160° W to 30° E (see Fig. 3). The moisture supplied to the environment is then estimated as  $M_{\text{LT, large}} = -\langle q d\omega/dp H(\Delta) H(-\omega_1) \rangle$ , where  $q$  is the specific humidity,  $\langle \dots \rangle$  indicates the mean over the restricted calculation region, and  $H$  is the step function. The index  $D$  is computed as  $D \equiv \langle \Delta H(\Delta) H(-\omega_1) \rangle / \langle -\omega_2 H(-\omega_2) \rangle$ .

Values of  $D$  and  $S$  are similar over ten years of data or one year, and are similar whether individual months or long-term means for each month of the year are used. These indices capture over 25% of the ECS variance even if computed from only a single month of data from each model. Thus, long records are unnecessary for deducing the strength of lower-tropospheric mixing.

The reason for restricting calculation of  $D$  to the cooler tropical longitudes is that a few climate models erroneously place much of the shallow ascent over warm oceans, where it does not seem to contribute as much to low-cloud feedback. In observations, and in most models, the restriction has little effect because most of the shallow ascent persistent enough to appear in monthly-mean data is already located in the specified region. We speculate that the location of the ascent matters because the associated shallow descent is more relevant if it occurs over, or upstream of, regions of radiatively important low cloud.

Both lower-tropospheric mixing indices retain statistically significant correlations with ECS for all alterations to their definitions that we tried. Specifically, the correlation of  $S$  with ECS ( $r_{S-\text{ECS}}$ ) is similar with  $\omega_{500}$  percentiles of 0.25 or 0.5, but drops with looser thresholds, which begin to pick up parts of the resolved lower-tropospheric mixing region. Tighter thresholds reduce the spread in  $S$  between models, reducing  $r_{S-\text{ECS}}$ . The correlation  $r_{D-\text{ECS}}$  is somewhat weaker (as low as 0.3) if the longitudinal restriction for  $D$  is removed, or if other definitions of  $\omega_1$  and  $\omega_2$  are used.

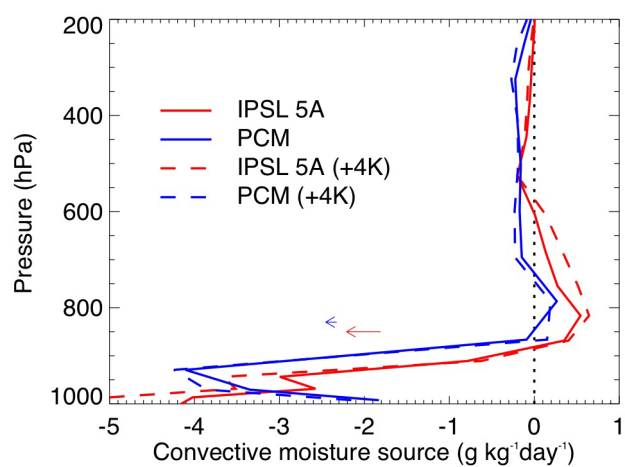
34. Sherwood, S. C., Meyer, M. L., Allen, R. J. & Titchner, H. A. Robust tropospheric warming revealed by iteratively homogenized radiosonde data. *J. Clim.* **21**, 5336–5352 (2008).
35. Sherwood, S. C., Roca, R., Weckwerth, T. M. & Andronova, N. G. Tropospheric water vapor, convection and climate. *Rev. Geophys.* **48**, RG2001 (2010).



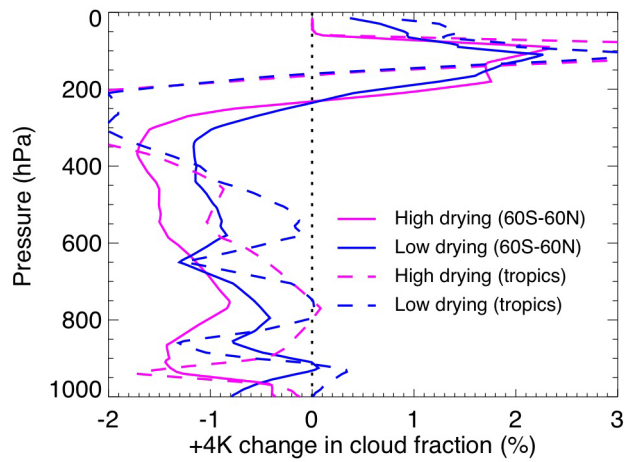
**Extended Data Figure 1 | Illustration of atmospheric overturning circulations.** Deep overturning strongly coupled to the hydrological cycle and atmospheric energy budget is shown by solid lines; lower-tropospheric mixing is shown by dashed lines. The MILC feedback results from the increasing

relative role of lower-tropospheric mixing in exporting humidity from the boundary layer as the climate warms, thus depleting the layer of water vapour needed to sustain low cloud cover.



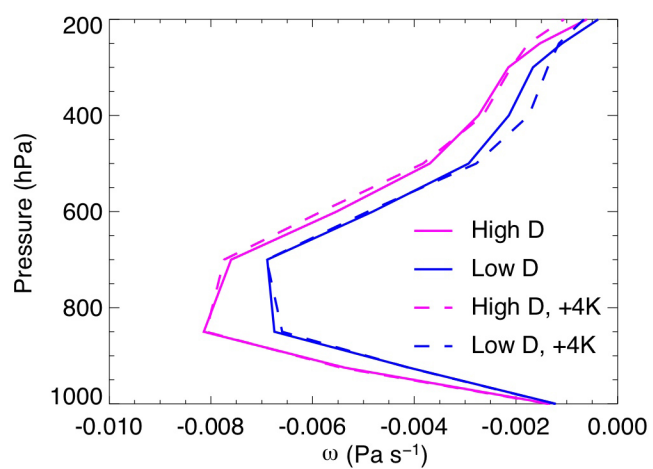


**Extended Data Figure 2 | Small-scale moisture source  $M_{\text{small}}$ .** Vertical profile averaged over all tropical oceans, for two selected climate models (see legend) with very different warming responses, in present-day (solid) and +4 K (dashed) climates.

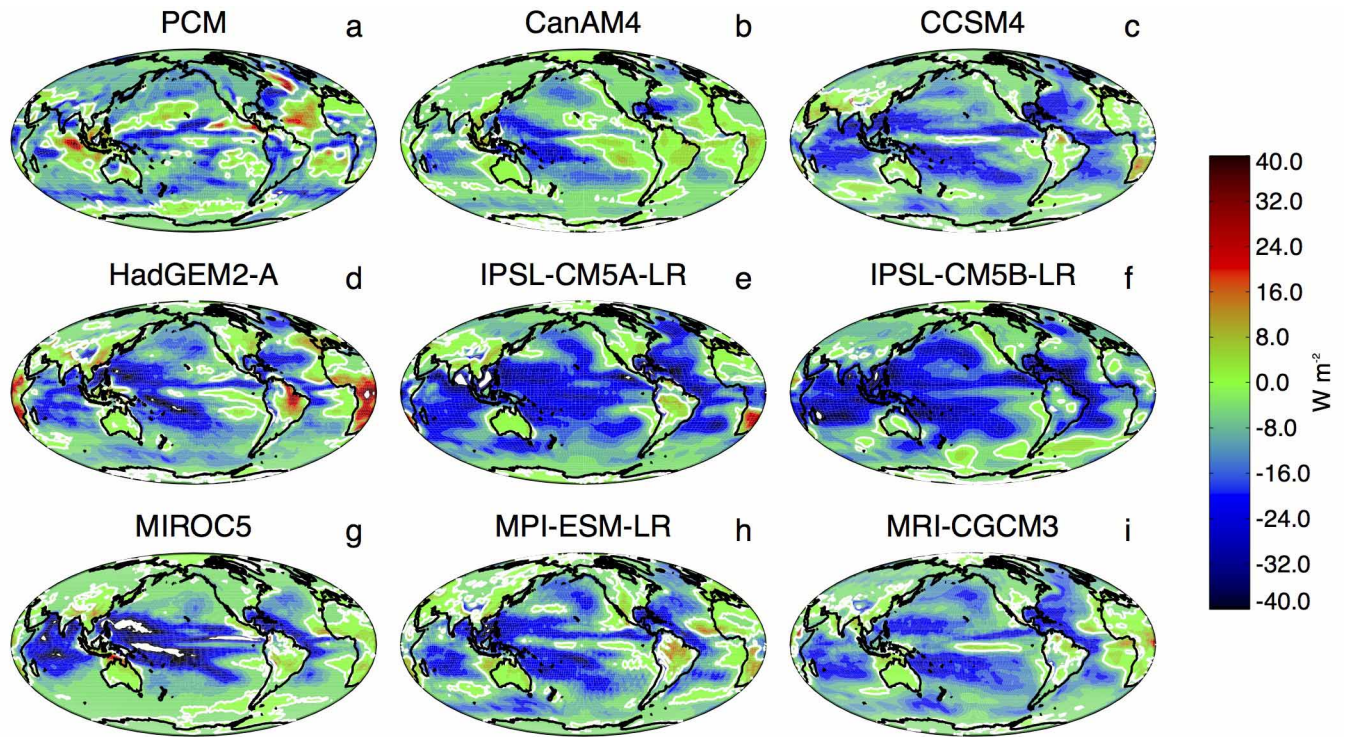


**Extended Data Figure 3 | Response of cloud fraction to warming.** Profile of average change in model cloud fractional cover at +4 K in the four atmosphere models with largest (magenta) and smallest (blue) estimated +4 K increases in planetary-boundary-layer drying, averaged from 30° S to 30° N (dashed) or 60° S to 60° N (solid). The drying estimate is obtained by adding the explicitly computed change in  $M_{LT, large}$  to the change in  $M_{sm}$  estimated from  $S$  via the relationship shown in Fig. 2a. The typical mean cloud fraction below 850 hPa is about 10% to 20%, and the changes shown are absolute changes in this fraction, so are of the order of 10% of the initial cloud cover.





**Extended Data Figure 4 | Response of large-scale lower-tropospheric mixing to warming.** Profiles of mean vertical velocity in regions of shallow ascent, in control and +4 K climates. The similarity of dashed and solid lines indicates that mass overturning associated with these regions is roughly the same in the warmer simulations, on average.



**Extended Data Figure 5 | Response of small-scale, low-level drying to warming.** Change in convective moisture source  $M_{\text{small}}$  below 850 hPa upon a +4 K warming in eight atmosphere models and one CMIP3 coupled model; units are  $\text{W m}^{-2}$ , with negative values indicating stronger drying near the

surface. Zero contours are shown in white (a few off-scale regions also appear white). The models used for calculating  $M_{\text{large}}$  are the eight shown here plus two for which  $M_{\text{small}}$  data were unavailable: CNRM-CM5 and FGOALS-g2.



Extended Data Table 1 | List of CMIP5 coupled models used

Model	Centre	Forcing ( $\text{W m}^{-2}$ )	Total feedback ( $\text{W m}^{-2} \text{K}^{-1}$ )	ECS (K)
ACCESS1-0	ACCESS	3.01	-0.79	3.79
ACCESS1-3	ACCESS	2.96	-0.86	3.45
BCC-CSM1-1	BCC	3.35	-1.16	2.88
BNU-ESM	GCESS/BNU	3.78	-0.92	4.11
CanESM2	CCC	3.85	-1.05	3.68
CCSM4	NCAR	3.70	-1.27	2.92
CESM1-BGC	NCAR	—	—	—
CESM1-CAM5	NCAR	—	—	—
CMCC-CM	CMCC	—	—	—
CNRM-CM5	CNRM	3.71	-1.14	3.25
CSIRO-Mk3-6-0	CSIRO/QCCCE	2.63	-0.66	3.99
FGOALS-g2	LASG/IAP	2.89	-0.84	3.45
FGOALS-s2	LASG/IAP	3.84	-0.92	4.16
GFDL-CM3	GFDL	3.00	-0.76	3.96
GFDL-ESM2G	GFDL	3.11	-1.31	2.38
GFDL-ESM2M	GFDL	3.41	-1.41	2.41
GISS-E2-H	GISS	3.83	-1.66	2.30
GISS-E2-R	GISS	3.77	-1.79	2.11
HadGEM2-ES	MOHC	2.95	-0.65	4.55
INMCM4	INM	2.98	-1.44	2.07
IPSL-CM5A-LR	IPSL	3.12	-0.76	4.10
IPSL-CM5B-LR	IPSL	2.66	-1.03	2.59
MIROC5	MIROC	4.16	-1.54	2.71
MIROC-ESM	MIROC	4.27	-0.92	4.65
MPI-ESM-LR	MPI	4.15	-1.15	3.60
MPI-ESM-MR	MPI	4.11	-1.20	3.44
MPI-ESM-P	MPI	4.35	-1.27	3.42
MRI-CGCM3	MRI	3.26	-1.26	2.59
NorESM1-ME	NCC	—	—	—
NorESM1-M	NCC	3.21	-1.13	2.83

Centre acronyms used to identify them in scatter plots are also shown. The derived forcing, total feedback, and equilibrium climate sensitivities are given for models with abrupt  $4 \times \text{CO}_2$  simulations.

Extended Data Table 2 | List of CMIP3 coupled models used

Model	Centre	ECS (K)
CCCMA-CGCM3 1	CCC	3.4
CCCMA-CGCM3 1 T63	CCC	3.4
GFDL-CM2-0	GFDL	2.9
GFDL-CM2-1	GFDL	3.4
GISS-MODEL-E-H	GISS	2.7
GISS-MODEL-E-R	GISS	2.7
IAP-FGOALS1-0-G	IAP	2.3
INGV-ECHAM4	INGV	—
INMCM3-0	INM	2.1
IPSL-CM4	IPSL	4.4
MIROC3-2-HIRES	MIROC	4.3
MIROC3-2-MEDRES	MIROC	4.0
MPI-ECHAM5	MPI	3.4
MRI-CGCM2-3-2A	MRI	3.2
NCAR-CCSM3-0	NCAR	2.7
NCAR-PCM1	NCAR	2.1
UKMO-HadCM3	MOHC	3.3
UKMO-HadGEM1	MOHC	4.4

Centre acronyms used to identify them in scatter plots are also shown, as are feedback values given by ref. 28.



# The complete genome sequence of a Neanderthal from the Altai Mountains

Kay Prüfer<sup>1</sup>, Fernando Racimo<sup>2</sup>, Nick Patterson<sup>3</sup>, Flora Jay<sup>2</sup>, Sriram Sankararaman<sup>3,4</sup>, Susanna Sawyer<sup>1</sup>, Anja Heinze<sup>1</sup>, Gabriel Renaud<sup>1</sup>, Peter H. Sudmant<sup>5</sup>, Cesare de Filippo<sup>1</sup>, Heng Li<sup>3</sup>, Swapan Mallick<sup>3,4</sup>, Michael Dannemann<sup>1</sup>, Qiaomei Fu<sup>1,6</sup>, Martin Kircher<sup>1,5</sup>, Martin Kuhlwilm<sup>1</sup>, Michael Lachmann<sup>1</sup>, Matthias Meyer<sup>1</sup>, Matthias Ongyerth<sup>1</sup>, Michael Siebauer<sup>1</sup>, Christoph Theunert<sup>1</sup>, Arti Tandon<sup>3,4</sup>, Priya Moorjani<sup>4</sup>, Joseph Pickrell<sup>4</sup>, James C. Mullikin<sup>7</sup>, Samuel H. Vohr<sup>8</sup>, Richard E. Green<sup>8</sup>, Ines Hellmann<sup>9</sup>, Philip L. F. Johnson<sup>10</sup>, Hélène Blanche<sup>11</sup>, Howard Cann<sup>11</sup>, Jacob O. Kitzman<sup>5</sup>, Jay Shendure<sup>5</sup>, Evan E. Eichler<sup>5,12</sup>, Ed S. Lein<sup>13</sup>, Trygve E. Bakken<sup>13</sup>, Liubov V. Golovanova<sup>14</sup>, Vladimir B. Doronichev<sup>14</sup>, Michael V. Shunkov<sup>15</sup>, Anatoli P. Derevianko<sup>15</sup>, Bence Viola<sup>16</sup>, Montgomery Slatkin<sup>2</sup>, David Reich<sup>3,4,17</sup>, Janet Kelso<sup>1</sup> & Svante Pääbo<sup>1</sup>

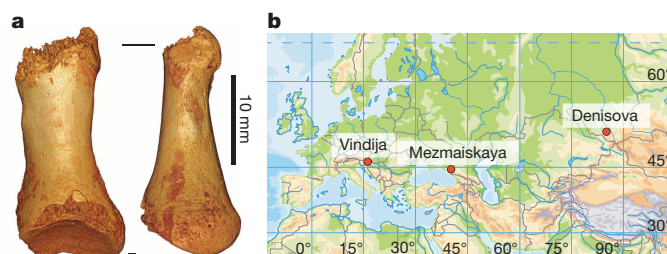
**We present a high-quality genome sequence of a Neanderthal woman from Siberia. We show that her parents were related at the level of half-siblings and that mating among close relatives was common among her recent ancestors. We also sequenced the genome of a Neanderthal from the Caucasus to low coverage. An analysis of the relationships and population history of available archaic genomes and 25 present-day human genomes shows that several gene flow events occurred among Neanderthals, Denisovans and early modern humans, possibly including gene flow into Denisovans from an unknown archaic group. Thus, interbreeding, albeit of low magnitude, occurred among many hominin groups in the Late Pleistocene. In addition, the high-quality Neanderthal genome allows us to establish a definitive list of substitutions that became fixed in modern humans after their separation from the ancestors of Neanderthals and Denisovans.**

In 2008, a hominin finger phalanx was discovered during excavation in the east gallery of Denisova Cave in the Altai Mountains. From this bone, a genome sequence was determined to ~30-fold coverage<sup>1</sup>. Analysis showed that it came from a previously unknown group of archaic humans related to Neanderthals which we named ‘Denisovans’<sup>2</sup>. Thus, at least two distinct human groups, Neanderthals and the related Denisovans, inhabited Eurasia when anatomically modern humans emerged from Africa. In 2010, another hominin bone, this time a proximal toe phalanx (Fig. 1a), was recovered in the east gallery of Denisova Cave<sup>3</sup>. Layer 11, where both the finger and the toe phalanx were found, is thought to be at least 50,000 years old. The finger was found in sublayer 11.2, which has an absolute date of  $50,300 \pm 2,200$  years (OxA-V-2359-16), whereas the toe derives from the lowest sublayer 11.4, and may thus be older than the finger (Supplementary Information sections 1 and 2a). The phalanx comes from the fourth or the fifth toe of an adult individual and its morphological traits link it with both Neanderthals and modern humans<sup>3</sup>.

## Genome sequencing

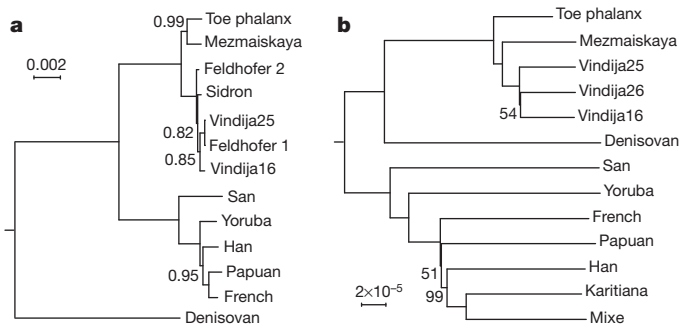
In initial experiments to determine if DNA was preserved in the toe phalanx, we extracted and sequenced random DNA fragments. This revealed that about 70% of the DNA fragments present in the specimen aligned to the human genome. Initial inspection of the fragments with similarity to the mitochondrial (mt) genome suggested that its mtDNA was closely related to Neanderthal mtDNAs. We therefore assembled the

full mitochondrial sequence by aligning DNA fragments to a complete Neanderthal mitochondrial genome<sup>4</sup> (Supplementary Information section 2b). A phylogenetic tree (Fig. 2a) shows that the toe phalanx mtDNA shares a common ancestor with six previously published Neanderthal mtDNAs<sup>5</sup> to the exclusion of present-day humans and the Denisova finger phalanx. Among Neanderthal mtDNAs, the toe mtDNA is most closely related to the mtDNA from infant 1 from Mezmaiskaya Cave in the Caucasus<sup>6</sup>.



**Figure 1 | Toe phalanx and location of Neanderthal samples for which genome-wide data are available. a**, The toe phalanx found in the east gallery of Denisova Cave in 2010. Dorsal view (left image), left view (right image). Total length of the bone is 26 mm. **b**, Map of Eurasia showing the location of Vindija Cave, Mezmaiskaya Cave and Denisova Cave, where Neanderthal samples used here were found.

<sup>1</sup>Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. <sup>2</sup>Department of Integrative Biology, University of California, Berkeley, California 94720-3140, USA. <sup>3</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. <sup>4</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>5</sup>Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. <sup>6</sup>Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China. <sup>7</sup>Genome Technology Branch and NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>8</sup>Department of Biomolecular Engineering, University of California, Santa Cruz, California 95064, USA. <sup>9</sup>Max F. Perutz Laboratories, Mathematics and Bioscience Group, Campus Vienna Biocenter 5, Vienna 1030, Austria. <sup>10</sup>Department of Biology, Emory University, Atlanta, Georgia 30322, USA. <sup>11</sup>Fondation Jean Dausset, Centre d'Étude du Polymorphisme Humain (CEPH), 75010 Paris, France. <sup>12</sup>Howard Hughes Medical Institute, Seattle, Washington 98195, USA. <sup>13</sup>Allen Institute for Brain Science, Seattle, Washington 98103, USA. <sup>14</sup>ANO Laboratory of Prehistory 14 Linia 3-11, St. Petersburg 1990 34, Russia. <sup>15</sup>Palaeolithic Department, Institute of Archaeology and Ethnography, Russian Academy of Sciences, Siberian Branch, 630090 Novosibirsk, Russia. <sup>16</sup>Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. <sup>17</sup>Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. †Present address: Ludwig-Maximilians-Universität München, Martinsried, 82152 Munich, Germany.



**Figure 2 | Phylogenetic relationships of the Altai Neanderthal.** **a**, Bayesian tree of mitochondrial sequences of the toe phalanx, the Denisovan finger phalanx, six Neanderthals and five present-day humans. Posterior probabilities are given for branches whose support is less than one (Supplementary Information section 2b). **b**, Neighbour-joining tree based on autosomal transversion differences among the toe phalanx, four Neanderthals, the Denisova genome and seven present-day human individuals. Bootstrap values are shown for branches supported by less than 100% of 1,000 bootstrap replicates (Supplementary Information section 6).

We generated four DNA libraries using a recently published protocol that is particularly efficient in retrieving DNA from ancient samples<sup>1,7</sup>. These libraries, together with one library prepared using a previous protocol<sup>8</sup>, were treated with uracil-DNA glycosylase to remove uracil residues, a common miscoding lesion in ancient DNA that results from the deamination of cytosine<sup>9–11</sup> (Supplementary Information section 5a). In total, these five DNA libraries provided 52-fold sequence coverage of the genome. We estimated present-day human DNA contamination in the libraries with four complementary approaches (Supplementary Information section 5) using mtDNA and nuclear DNA and conclude that present-day human contamination among the DNA fragments sequenced is around 1%. After genotype calling, which is designed to be insensitive to low levels of error, we expect that the inferred genome sequence is largely free from contamination.

### Relationship to other hominins

We compared the toe phalanx genome to the Denisovan genome<sup>1</sup>, the draft Neanderthal genome of 1.3-fold coverage determined from three individuals from Vindija Cave, Croatia<sup>12</sup>, the genome of a Neanderthal infant estimated to be 60,000 to 70,000-years-old<sup>13</sup> from Mezmaiskaya Cave in the Caucasus that we sequenced to 0.5-fold genomic coverage (Supplementary Information section 1; Fig. 1b) as well as 25 genomes of present-day humans: 11 previously sequenced to between 24- and 31-fold coverage<sup>1</sup> (Panel A), and 14 sequenced to between 35- and 42-fold coverage for this study (Panel B). We used pooled fosmid sequencing to resolve the sequences of the two chromosomes carried by 13 of these individuals<sup>14</sup> (Supplementary Information section 4).

A neighbour-joining tree (Fig. 2b) based on transversions, that is, purine–pyrimidine differences, among 7 present-day humans and the low-coverage Mezmaiskaya and Vindija genomes corrected for errors (Supplementary Information section 6a), shows that the toe phalanx nuclear genome forms a clade with the genomes of Neanderthals. The average DNA sequence divergence between the toe phalanx genome and the Mezmaiskaya and Vindija Neanderthal genomes is approximately a third of that between the Neanderthal and Denisova genomes. We conclude that the individual from whom the toe phalanx derives is a Neanderthal. Hereafter we refer to it as the ‘Altai Neanderthal’.

### Branch shortening

The length of the branches leading from the common ancestor shared with chimpanzee to the high-coverage Altai Neanderthal and Denisovan genomes are 1.02% (range of point estimates: 0.99–1.05%) and 0.81% (range: 0.77–0.84%) shorter, respectively, than the branches to the present-day human genomes (Table 1; Supplementary Information section 6b). This is expected because the archaic genomes ceased accumulating substitutions at the death of the individual tens of thousands of years ago. We previously estimated the shortening of the Denisovan lineage to be 1.16% (range: 1.13–1.27%)<sup>1</sup>. The fact that using present-day human genomes of higher quality and more stringent quality filtering reduces the estimate by about a third shows that at present, estimates of lineage lengths are unstable, probably owing to differences in the error rates among the genomes used. Nevertheless, the fact that the Neanderthal lineage is about 20% shorter than the Denisovan lineage suggests that the Neanderthal toe phalanx is older than the Denisovan finger phalanx, consistent with the stratigraphy of the cave.

### Population split times

Figure 2b reflects the average divergence between DNA sequences. The times when the ancestral populations of archaic and modern humans separated are by necessity earlier. We used two approaches to estimate these population split times (Supplementary Information section 12). We caution that for these and other age estimates we rely on dates for the divergence of human and chimpanzee DNA sequences that in turn depend on the human mutation rate, which is currently controversial. In the text we present estimates based on a mutation rate of  $0.5 \times 10^{-9}$  base pairs per year, estimated from comparisons of the genomes of parents and children<sup>15–19</sup>. In Table 1 we also present estimates based on a rate of  $1.0 \times 10^{-9}$  base pairs per year derived from the fossil record which was used in previous studies of archaic genomes<sup>1,2,12</sup>. We also caution that the split times are at the best approximate because the models of population history used are likely to be inaccurate.

We first estimated population split times by extending the pairwise sequentially Markovian coalescent model (PSMC) to estimate the distribution of coalescence times between two single chromosomes that come from different populations<sup>20,21</sup> (Supplementary Information section 12). Using Sub-Saharan African genomes that were experimentally

**Table 1 | Dating for branch shortening and population splits**

Event	As % of human–chimp divergence	Absolute date calibration number 1 in kyr ( $\mu = 1 \times 10^{-9}$ per bp per year)	Absolute date calibration number 2 in kyr ( $\mu = 0.5 \times 10^{-9}$ per bp per year)	Supplementary Information section
Altai Neanderthal branch shortening	0.99–1.05	64–68	129–136	6b
Denisova branch shortening	0.77–0.84	50–54	100–109	6b
San–West African split	0.66–1.00	43–65	86–130	12
Introgressing Neanderthal–Altai split	0.58–0.88	38–57	77–114	13
Introgressing Denisovan–Denisovan split	2.12–3.10	138–202	276–403	13
Neanderthal–Denisova split*	2.93–3.64	190–236	381–473	12
Archaic–African split*	4.23–5.89	275–383	550–765	12
Unknown archaic split	7.90–31.12	450–2027	900–4054	16a, b

This table gives date ranges for two calibrations. The first assumes human–chimpanzee divergence of 6.5 million years and 1.30% for human–chimp divergence, or a mutation rate of  $1 \times 10^{-9}$  bp per year<sup>1,2,12</sup>. The second is based on direct measurement of per generation mutation rates<sup>15–17</sup>, corresponding to a mutation rate of  $0.5 \times 10^{-9}$  per bp per year or 13 million years ago for human–chimpanzee divergence, and may fit better with some aspects of the fossil record<sup>45,46</sup>. Intervals give the range of values over tested human genomes for branch shortening; lowest and highest estimate for two or three methods for San–West African, Neanderthal–Denisova, Neanderthal–African and Denisova–African split; jackknife confidence interval over introgressed chunks for the Introgressing–Archaic–African splits; and a union of the jackknife confidence interval in Supplementary Information section 16a and the highest posterior mode in Supplementary Information section 16b for the unknown archaic split.

\*The indicated values are corrected for branch shortening where relevant as described in the Supplementary Information.

phased (Supplementary Information section 14) and segments of the archaic genomes in which the two chromosomes within an individual are closely related, we estimate the population split time between modern humans on the one hand, and Neanderthals and Denisovans on the other, to between 553,000 and 589,000 years ago, and the split time between Neanderthals and Denisovans to 381,000 years ago (Supplementary Information section 12).

In a second approach we counted how often randomly chosen alleles in an individual from one population are derived (that is, different from the apes) at positions where both the derived and ancestral alleles are seen in an individual from a second population<sup>1,12</sup>. Such derived alleles will be less frequent the older the population separation time is because more derived alleles in the second population will then be because of mutations that occurred after the split. Using this approach and the demographic history inferred from the PSMC (Supplementary Information section 12), we estimate the population split of Neanderthals and Denisovans from modern humans to 550,000–765,000 years ago, and the split time of Neanderthals and Denisovans to 445,000–473,000 years ago.

## Inbreeding

We noticed that the Altai Neanderthal genome contains several long runs of homozygosity, indicating that her parents were closely related (Fig. 3a). To estimate the extent of their relatedness, we scanned the genome for 1 Mb regions where most non-overlapping 50-kb windows were devoid of heterozygous sites and merged adjacent regions (Supplementary Information section 10). The Neanderthal genome has 20

such regions longer than 10 cM, whereas the Denisovan genome has one. We performed simulations of inbreeding scenarios that can result in regions of this number and length, and find that the inbreeding coefficient is 1/8, indicating that the parents were as closely related as half-siblings. As the Altai individuals is a female (Supplementary Information section 5) and the X chromosome also has long runs of homozygosity, we can exclude parental relationships in which none or only one of the two X chromosomes was inherited from closely-related common ancestor(s), that is, scenarios that include two successive males in the pedigree. We conclude that the parents of this Neanderthal individual were either half-siblings who had a mother in common, double first cousins, an uncle and a niece, an aunt and a nephew, a grandfather and a granddaughter, or a grandmother and a grandson (Fig. 3b).

To investigate whether mating between closely related individuals may have been typical of the Altai Neanderthal population, we examined the distribution of runs of homozygosity between 2.5 and 10 cM in length. After removing the runs expected from recent inbreeding, the Altai Neanderthal genome still contains more runs than the Denisovan genome ( $P < 2.2 \times 10^{-16}$ ), and both archaic genomes contain more than the Karitiana, a present-day population known to have a small effective size<sup>22</sup> (Fig. 3c; Supplementary Information 10). The sequencing of additional Neanderthal genomes to high quality will address whether breeding among close relatives was common also among Neanderthals in other geographic areas.

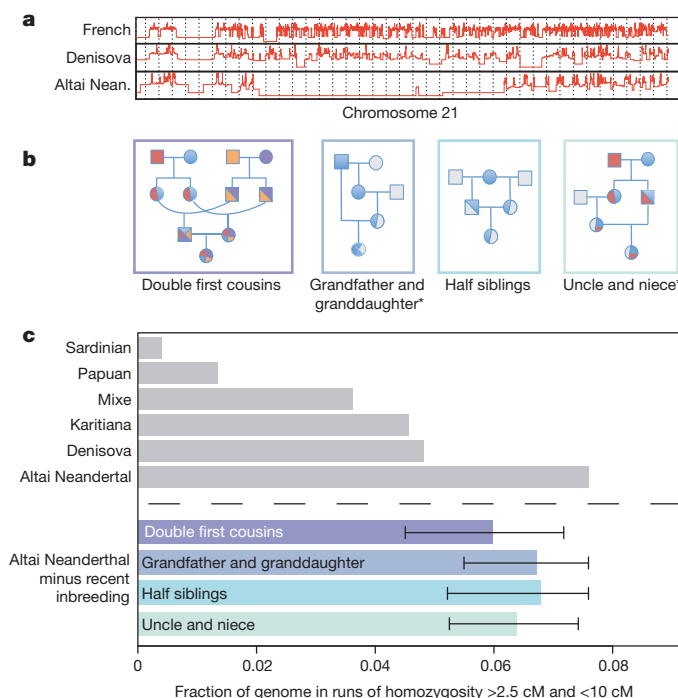
## Heterozygosity and population size

The Neanderthal autosomal genome carries 1.7–1.8 heterozygous sites per 10,000 bp (Supplementary Information section 9). This is 84% of the number of heterozygous sites in the Denisovan genome, 22–30% of that in present-day non-African genomes, and 16–18% of that in present-day African genomes (Extended Data Fig. 1). When regions of homozygosity longer than 2.5 cM stemming from recent as well as long-term inbreeding in the Neanderthal are removed, 2.1–2.2 sites per 10,000 are heterozygous, similar to what is observed in the Denisovan genome. Thus, heterozygosity in Neanderthals as well as Denisovans appears to have been lower than in present-day humans and is among the lowest measured for any organism<sup>23</sup>.

The demographic history of the population can be reconstructed from the distribution of the times since the most recent common ancestor of the two copies of the genome that a single person carries. We use the PSMC<sup>20</sup> to infer changes in the size of the Neanderthal population over time and compare this to inferences from the Denisovan and present-day human genomes (Fig. 4) (Supplementary Information section 12). All genomes analysed show evidence of a reduction in population size that occurred sometime before 1.0 million years ago. Subsequently, the population ancestral to present-day humans increased in size, whereas the Altai and Denisovan ancestral populations decreased further in size. It is thus clear that the demographic histories of both archaic populations differ substantially from that of present-day humans.

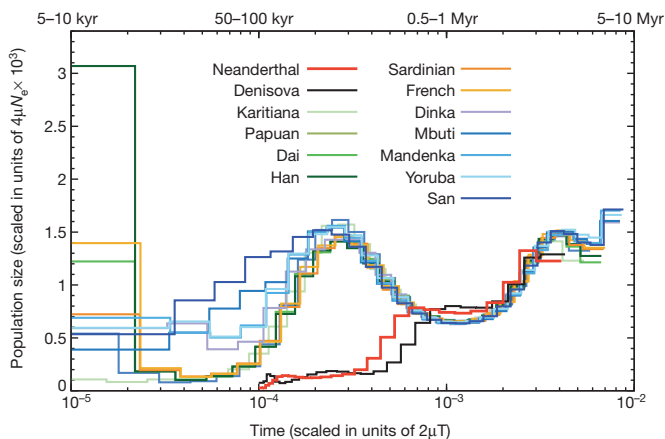
## Neanderthal gene flow into modern humans

We have previously shown that Neanderthals contributed parts of their genomes to present-day populations outside Africa<sup>12</sup> and that Denisovans contributed to the genomes of present-day populations in Oceania<sup>2,24</sup> (used here to refer to Australia, Melanesia and the Philippines). Using the high-coverage Neanderthal genome in conjunction with the two other Neanderthal genomes, we now estimate that the proportion of Neanderthal-derived DNA in people outside Africa is 1.5–2.1% (Supplementary Information 14; Extended Data Table 1). Second, we find that the Neanderthal-derived DNA in all non-Africans is more closely related to the Mezmaiskaya Neanderthal from the Caucasus than it is to either the Neanderthal from Siberia (Extended Data Table 2; Supplementary Information section 14) (Z-score range: 4.0–6.4) or to the Vindija Neanderthals from Croatia<sup>12</sup> (Z-score range: 1.7–3.9). These results cannot be explained by present-day human contamination in



**Figure 3 | Indications of inbreeding in the Altai Neanderthal individual.** **a**, Time since the most recent common ancestor in log-scale for the two alleles of a French, the Denisovan and the Altai Neanderthal individual (Supplementary Information section 12) along 40 Mb of chromosome 21. **b**, Pedigrees showing four possible scenarios of parental relatedness for the Altai Neanderthal (that is, the child at the bottom of each pedigree). Two additional scenarios can be derived by switching the sex of the parents for the panels marked with an asterisk. **c**, Fraction of the genome in runs of homozygosity between 2.5 and 10 cM in length for Altai Neanderthal, Denisovan and the three present-day human individuals with the largest fractions (grey bars). The fractions for the Altai Neanderthal (bottom four bars) are reduced by the fraction expected from the four inbreeding scenarios in **b**. Error bars represent the full range of values obtained from 700 simulations for each scenario.





**Figure 4 | Inference of population size change over time.** The y axis specifies a number proportional to the population size  $N_e$ . The x axis specifies time in units of divergence per base pair (along the top in years for mutation rates of  $0.5 \times 10^{-9}$  to  $1.0 \times 10^{-9}$  per site per year). The analysis assumes that the Neanderthal and Denisova remains are of the same age, whereas archaeological evidence and the branch shortening indicate that the Neanderthal bone is older than the Denisovan bone. However, because the exact difference in ages is not known, it is not possible to determine whether the reduction in population size experienced by both archaic groups (but not by modern humans) coincided in time.

the Mezmaiskaya Neanderthal data, as a contamination level on the order of 2.0–5.4% would be needed to account for the excess relatedness to the Mezmaiskaya Neanderthal whereas the contamination in the Mezmaiskaya data are estimated to be 0–1.1% (Supplementary Information 5a).

### Denisovan gene flow in mainland Asia

We used the two high-coverage archaic genomes and a hidden Markov model (HMM) to identify regions of specifically Neanderthal and specifically Denisovan ancestry in 13 experimentally phased present-day human genomes<sup>1,14</sup> (Supplementary Information sections 4 and 13). In the Sardinian and French genomes from Europe we find genomic regions of Neanderthal origin and few or no regions of Denisovan origin. In contrast, in the Han Chinese, the Dai in southern China, and the Karitiana and Mixe in the Americas, we find, in addition to regions of Neanderthal origin, regions that are consistent with being of Denisovan origin (Z-score = 4.3 excess relative to the Europeans) (Supplementary Information section 13), in agreement with previous analysis based on low-coverage archaic genomes<sup>25</sup>. These regions are also more closely related to the Denisova genome than the few regions identified in Europeans (Supplementary Information section 13). We estimate that the Denisovan contribution to mainland Asian and Native American populations is ~0.2% and thus about 25 times smaller than the Denisovan contribution to populations in Papua New Guinea and Australia. The failure to detect any larger Denisovan contribution in the genome of a 40,000-year-old modern human from the Beijing area<sup>26</sup> suggests that any Denisovan contribution to modern humans in mainland Asia was always quantitatively small. In fact, we cannot, at the moment, exclude that the Denisovan contribution to people across mainland Asia is owing to gene flow from ancestors of present-day people in Oceania after they mixed with Denisovans. We also note that in addition to this Denisovan contribution, the genomes of the populations in Asia and America appear to contain more regions of Neanderthal origin than populations in Europe<sup>1,27</sup> (Supplementary Information sections 13 and 14).

### Archaic population differentiation

To estimate how closely related the archaic populations that contributed DNA to present-day humans were to the archaic individuals from which high-coverage genomes have been determined, we compared

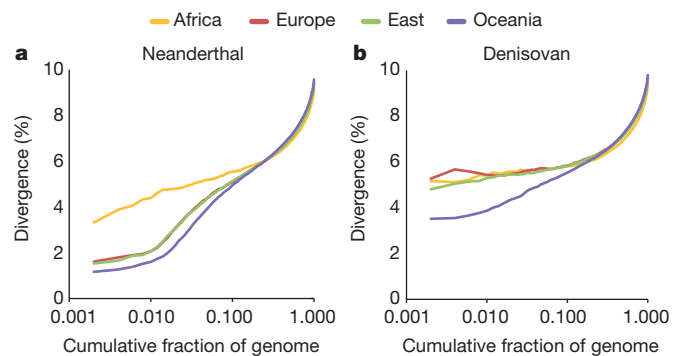
the regions of Neanderthal and Denisovan ancestry in present-day human genomes identified by an HMM to the sequenced archaic genomes (Supplementary Information section 13). We find that the DNA sequence divergence in the regions that are most similar between the Altai Neanderthal genome and the Neanderthals that contributed DNA to present-day Eurasians is ~1.35% of the human–chimpanzee divergence, whereas the regions with the smallest sequence divergence between the Denisovan genome and the Denisovans that contributed DNA to present-day Papuans and Australians is ~3.18%. Regions of similarly low divergence are also identified by a window-based comparison (Fig. 5).

We estimate the population split time between the introgressing Neanderthal and the Altai Neanderthal genome to 77,000–114,000 years ago, and the split time between the introgressing Denisovan and the Denisovan genome to 276,000–403,000 years ago (Supplementary Information section 13) (Table 1). This is consistent with the Denisovan population being larger, more diverse and/or more subdivided than Neanderthal populations, and with the idea that Denisovans may have populated a wide geographical area. It is also in agreement with the low diversity among Neanderthal nuclear<sup>2</sup> and mitochondrial<sup>5</sup> genomes.

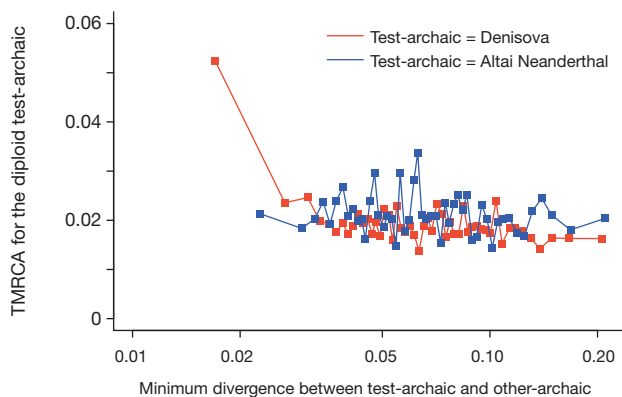
### Neanderthal gene flow into Denisovans

If gene flow occurred between Neanderthals and Denisovans, we would expect that regions of the genome where the divergence between Denisovan and Neanderthal haplotypes is low would carry many differences between the two haplotypes of the individual who harbours the introgressed genetic material. This is because this individual carries two haplotypes that have accumulated differences independently in the two populations. In contrast, in the absence of gene flow, regions of low divergence between a Neanderthal and a Denisovan haplotype are not expected to have particularly elevated diversity (Supplementary Information section 15).

We plotted the number of differences between the Neanderthal genome and the closest inferred DNA sequences in the Denisovan genome against Denisovan heterozygosity (Fig. 6). We find that Denisovan heterozygosity is increased in regions where the Neanderthal and one Denisovan allele are close, indicating that gene flow from Neanderthals into Denisovans occurred, and estimate that a minimum of 0.5% of the Denisovan genome was contributed by Neanderthals. The Denisovan genome shares more derived alleles with the Altai Neanderthal genome than with the Croatian or Caucasus Neanderthal genomes (Z-score range: 5.6–10.2) (Extended Data Table 2; Supplementary Information section 15), suggesting that the gene flow into Denisovans came from a



**Figure 5 | Relatedness of introgressing archaic and sequenced archaic samples.** Divergence of phased present-day human genomes to archaic genomes in windows of size 0.01 cM with a minimum of 25,000 analysed bases. Windows are sorted by sequence divergence measured on the archaic side of the tree (Supplementary Information section 13) and the y axis reports the divergence relative to human–chimpanzee divergence for cumulative fractions of the sorted windows over the entire genomes. Regions of low divergence between non-Africans and Neanderthals (a) and between Oceanians and Denisovans (b) indicate gene flow between these groups and the relative divergences between the introgressing archaic and sequenced archaic samples.



**Figure 6 | Neanderthal gene flow into Siberian Denisovans.** Divergence in 0.01 cM sized windows with at least 50 kb analysed bases between a 'test'-archaic genome and effectively haploid regions of the other archaic genome plotted against the most recent common ancestor of the two alleles of the 'test'-archaic. The plot shows 50 equally sized bins of windows for the 'test' Denisovan against the effectively haploid Neanderthal (red) and for the 'test'-archaic Altai Neanderthal against the effectively haploid Denisovan (blue). Divergence is given as percentage of human–chimpanzee divergence. Windows that show a close relationship between the effective haploid Altai Neanderthal and the closest inferred Denisovan haplotype show a deep divergence to the second Denisovan haplotype, indicating gene flow from Neanderthal into Denisovan.

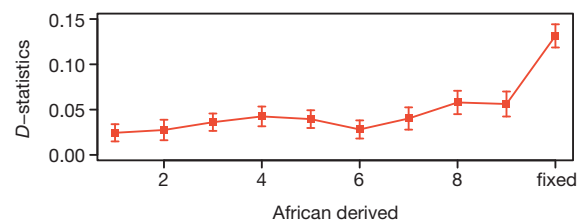
Neanderthal population more related to the Altai Neanderthal than to the other two Neanderthals. In the reciprocal analysis, we find no corresponding increase in Neanderthal heterozygosity.

Particularly strong signals of Neanderthal gene flow into Denisovans are found in the human leucocyte antigen (HLA) region and the CRISP gene cluster on chromosome 6 (Extended Data Fig. 2), where we find many segments for which one of the Denisova haplotypes and the Altai Neanderthal share a common ancestor within a few tens of thousands of years before the death of the Altai individual (Supplementary Information section 15). This suggests the possibility that introgressed Neanderthal alleles may have contributed to the Denisovan functional variation at the HLA and the CRISP cluster, which are involved in immunity and sperm function, respectively. This is interesting as it has been suggested that HLA alleles from Neanderthals and Denisovans have been of functional relevance in modern humans<sup>28</sup>.

### Unknown archaic gene flow into Denisovans

As the ancestors of both Neanderthals and Denisovans left Africa before the emergence of modern humans, one might expect present-day Africans to share equal proportions of derived alleles with these two archaic groups. However, we find that African genomes share about 7% more derived alleles with the Neanderthal genome than with the Denisova genome ( $Z = 11.6$  to  $13.0$ ; Extended Data Table 2; Supplementary Information section 16a) and that this is particularly the case for derived alleles that are fixed in Africans, of which 13–16% more are shared with the Neanderthal than with the Denisovan genome (Fig. 7).

We tested three non-mutually exclusive scenarios that could explain these observations. First, gene flow from the ancestor of Neanderthals after the split from Denisovans into the ancestors of all present-day humans would result in more sharing of derived alleles between present-day Africans and Neanderthals. However, because gene flow contributes alleles at low frequency, the sharing of derived alleles with Neanderthals would grow weaker with higher African derived allele frequency (Supplementary Information section 16a), whereas we observe the opposite (Fig. 7). Second, gene flow from the ancestors of present-day humans to Neanderthals after their split from Denisovans would also result in more sharing of derived alleles. However, the amount of allele frequency change (genetic drift) that has occurred in present-day Africans since the split from Neanderthals is too small to explain the extent of sharing



**Figure 7 | Altai and Denisovan allele sharing with Africans stratified by African allele frequency.** The plot shows the  $D$ -statistic of the form  $D$  (Neanderthal, Denisova; Africa, chimpanzee) binned by derived allele count in 10 deeply sequenced African genomes. Error bars represent  $\pm 1$  standard error. High-frequency and fixed derived alleles in Africa are more often shared with the Neanderthal than with the Denisovan genome.

of derived alleles fixed in Africans (Supplementary Information section 16a). Third, we considered a scenario where Denisovans received gene flow from a hominin whose ancestors diverged deeply from the lineage leading to Neanderthals, Denisovans and present-day humans. We find that this scenario is consistent with the data, as also suggested by others<sup>29</sup>, and estimate that 2.7–5.8% (jackknife 95% confidence interval) of the Denisova genome comes from this putative archaic hominin which diverged from the other hominins 0.9–1.4 million years ago (Supplementary Information section 16a). An approximate Bayesian computation<sup>30</sup> again supports the third scenario (Supplementary Information section 16b) and estimates that 0.5–8% of the Denisovan genome comes from an unknown hominin which split from other hominins between 1.1 and 4 million years ago.

We caution that these analyses make several simplifying assumptions. Despite these limitations, we show that the Denisova genome harbours a component that derives from a population that lived before the separation of Neanderthals, Denisovans and modern humans. This component may be present due to gene flow, or to a more complex population history such as ancient population structure maintaining a larger proportion of ancestral alleles in the ancestors of Denisovans over hundreds of thousands of years.

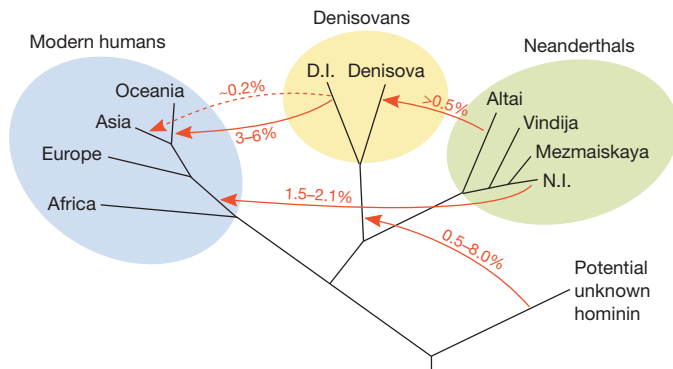
The putative admixture into Denisovans from an unknown archaic group raises the possibility that the apparent Denisovan contribution to the genomes of Papuans and Australians could originate from admixture with the same unknown archaic population instead of with Denisovans. However, we tested this hypothesis and found that the archaic component in the genomes of people in Papua New Guinea and Australia comes from a group related to the Denisovans and not from an unknown archaic hominin (Supplementary Information section 17).

### Copy number differences

The high-quality archaic genomes allow us to identify genetic changes that may have been relevant for putative biological traits that set modern humans apart from archaic humans. To identify genomic regions that have changed in copy number during hominin evolution, we used the variation of coverage along the two archaic genomes and 25 present-day human genomes (Supplementary Information section 8). We find three regions that have been duplicated only on the modern human lineage (Extended Data Table 3). One region overlaps *BOLA2*, which occurs as a single copy per haploid genome in the archaic genomes but has two to five copies in all but one of 675 present-day humans analysed, and which is near a microdeletion associated with developmental delay, intellectual disability and autism<sup>31</sup>.

### Catalogue of modern human changes

We compiled a genome-wide catalogue of sites where all or nearly all of 1,094 present-day humans<sup>32</sup> carry the same nucleotide but differ from the Neanderthal, Denisovan and great ape genomes (Supplementary Information section 18). In the regions of the genome to which short fragments can be mapped, there are 31,389 such single nucleotide



**Figure 8 | A possible model of gene flow events in the Late Pleistocene.** The direction and estimated magnitude of inferred gene flow events are shown. Branch lengths and timing of gene flows are not drawn to scale. The dashed line indicates that it is uncertain if Denisovan gene flow into modern humans in mainland Asia occurred directly or via Oceania. D.I. denotes the introgressing Denisovan, N.I. the introgressing Neanderthal. Note that the age of the archaic genomes precludes detection of gene-flow from modern humans into the archaic hominins.

substitutions and 4,113 short insertions and deletions (indels) shared by all present-day humans analysed, and a further 105,757 substitutions and 3,900 indels shared by 90% of present-day humans. This list of simple DNA sequence changes that distinguish modern humans from our nearest extinct relatives is thus comparatively small. For example, it contains only 96 fixed amino acid substitutions in a total of 87 proteins and in the order of three thousand fixed changes that potentially influence gene expression in present-day humans (Supplementary Information section 18).

Because the manner in which modern and archaic humans may have differed in aspects of their cognition is particularly interesting, we focused on the expression in the developing human brain of transcripts encoding the 87 proteins with fixed amino acid changes (Supplementary Information section 20). In comparison to a control set of transcripts that carry 108 silent substitutions fixed in present-day humans, genes carrying fixed amino acid changes are more often expressed in the ventricular zone of the developing neocortex ( $P = 0.06$ , corrected for multiple testing). Out of the five genes which are expressed in the proliferative layers (ventricular and subventricular zones combined) during mid-fetal development (*CASC5*, *KIF18A*, *TKTL1*, *SPAG5*, *VCAM1*), three (*CASC5*, *KIF18A*, *SPAG5*) are associated with the kinetochore of the mitotic spindle. This may be relevant phenotypically as the orientation of the mitotic cleavage plane in neural precursor cells during cortex development is thought to influence the fate of the daughter cells and the number of neurons generated (for example, see ref. 33). Another of these five genes, *VCAM1*, is essential for maintenance of neural stem cells in the adult subventricular zone<sup>34</sup>.

Another way to prioritize changes in the catalogue for functional studies is to identify those that show signs of having risen to high frequency rapidly as they may have been affected by positive selection. We implemented an HMM to scan the genome for regions where the Neanderthal and Denisovan genomes fall outside of the variation of present-day humans (Supplementary Information section 19a). We ranked these regions, which cover less than 100 Mb of the genome, according to genetic length, because regions that rose rapidly to fixation are expected to be longer as they have been less affected by recombination events. A set of 63 regions likely to have been affected by positive selection were identified (Supplementary Information Table S19a.3). They contain 2,123 substitutions and 61 indels that are fixed or of high-frequency (>90%) in modern humans (Supplementary Information section 19b). They include, for example, the gene *RB1CC1* (also called *FIP200*) which encodes a transcription factor which, like *VCAM1*, is essential for maintenance of neuronal stem cells in the adult subventricular zone<sup>35</sup>. In present-day humans, but not Neanderthals and Denisovans, *RB1CC1*

carries a substitution inferred to change an amino acid in the encoded protein as well as a substitution that affects a conserved site in a motif that occurs across the genome<sup>36</sup>. Functional investigations will be necessary to clarify whether these and other such changes affect any phenotypes in present-day humans.

## Discussion

We present evidence for three to five cases of interbreeding among four distinct hominin populations (Fig. 8). Clearly the real population history is likely to have been even more complex. For example, most cases of gene flow are likely to have occurred intermittently, often in both directions and across a geographic range. Thus, combinations of gene flow among different groups and substructured populations may have yielded the patterns detected rather than the discrete events considered here. Nevertheless, our analyses show that hominin groups met and had offspring on many occasions in the Late Pleistocene, but that the extent of gene flow between the groups was generally low.

We note that the observation that the Neanderthal DNA sequences in non-Africans share more derived alleles with the Neanderthal from the Caucasus than with Neanderthals from either Croatia or the Altai indicates that the archaic gene flow into non-Africans occurred at a time when Neanderthal populations had separated from each other. We also note that the introgressed Neanderthal DNA sequences suggest a population split from the Altai Neanderthal between 77,000 and 114,000 years ago (Supplementary Information section 13), well after ~230,000 years ago when Neanderthal features appear in the fossil record<sup>37</sup>. These and other results<sup>38,39</sup> show that the allele sharing between Neanderthals and non-African populations is owing to recent admixture rather than ancient population subdivision, an alternative which we and others previously considered possible<sup>12,40</sup>.

The evidence suggestive of gene flow into Denisovans from an unknown hominin is interesting. The estimated age of 0.9 to 4 million years for the population split of this unknown hominin from the modern human lineage is compatible with a model where this unknown hominin contributed its mtDNA to Denisovans since the Denisovan mtDNA diverged from the mtDNA of the other hominins about 0.7–1.3 million years ago<sup>41</sup>. The estimated population split time is also compatible with the possibility that this unknown hominin was what is known from the fossil record as *Homo erectus*. This group started to spread out of Africa around 1.8 million years ago<sup>42</sup>, but Asian and African *H. erectus* populations may have become finally separated only about one million years ago<sup>43</sup>. However, further work is necessary to establish if and how this gene flow event occurred.

## METHODS SUMMARY

Sequences were generated on the Illumina HiSeq 2500 and base-calling was carried out using Ibis<sup>44</sup>. For all present-day human samples there is informed consent consistent with their use for whole genome sequencing and dissemination of data. Reads were merged and adaptor trimmed as described<sup>1</sup> and mapped to the human reference genome using BWA (version 0.5.10). Genotyping was carried out using GATK (version 1.3). We restricted analyses to regions of the genome that are non-repetitive (excluding tandem repeats), unique (requiring at least 50%, or all, overlapping 35-mers covering a position to map uniquely, allowing for one mismatch), and fall within the central 95% of the coverage distribution corrected for GC bias (Supplementary Information section 5b). The Supplementary Information describes the details of data processing and other analyses.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 5 September; accepted 15 November 2013.**

**Published online 18 December 2013.**

1. Meyer, M. et al. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
2. Reich, D. et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
3. Mednikova, M. B. A proximal pedal phalanx of a paleolithic hominin from Denisova cave, Altai. *Archaeol. Ethnol. Anthropol. Eurasia* **39**, 129–138 (2011).



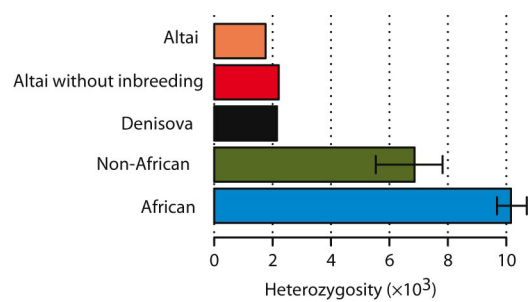
4. Green, R. E. *et al.* A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell* **134**, 416–426 (2008).
5. Briggs, A. W. *et al.* Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318–321 (2009).
6. Golovanova, L. V., Hoffercker, J. F., Kharitonov, V. M. & Romanova, G. P. Mezmaiskaya cave: A Neanderthal occupation in the Northern Caucasus. *Curr. Anthropol.* **40**, 77–86 (1999).
7. Gansauge, M. T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* **8**, 737–748 (2013).
8. Kircher, M. Analysis of high-throughput ancient DNA sequencing data. *Methods Mol. Biol.* **840**, 197–228 (2012).
9. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl Acad. Sci. USA* **104**, 14616–14621 (2007).
10. Briggs, A. W. *et al.* Removal of deaminated cytosines and detection of *in vivo* methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
11. Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A. & Paabo, S. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res.* **29**, 4793–4799 (2001).
12. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
13. Skinner, A. R. *et al.* ESR dating at Mezmaiskaya Cave, Russia. *Appl. Radiat. Isot.* **62**, 219–224 (2005).
14. Kitzman, J. O. *et al.* Haplotype-resolved genome sequencing of a Gujarati Indian individual. *Nature Biotechnol.* **29**, 59–63 (2011).
15. Abecasis, G. R. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
16. Awadalla, P. *et al.* Direct measure of the *de novo* mutation rate in autism and schizophrenia cohorts. *Am. J. Hum. Genet.* **87**, 316–324 (2010).
17. Roach, J. C. *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636–639 (2010).
18. Kong, A. *et al.* Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475 (2012).
19. Campbell, C. D. *et al.* Estimating the human mutation rate using autozygosity in a founder population. *Nature Genet.* **44**, 1277–1281 (2012).
20. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
21. Prado-Martinez, J. *et al.* Great ape genetic diversity and population history. *Nature* **499**, 471–475 (2013).
22. Kirin, M. *et al.* Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* **5**, e13996 (2010).
23. Leffler, E. M. *et al.* Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol.* **10**, e1001388 (2012).
24. Reich, D. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
25. Skoglund, P. & Jakobsson, M. Archaic human ancestry in East Asia. *Proc. Natl Acad. Sci. USA* **108**, 18301–18306 (2011).
26. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
27. Wall, J. D. *et al.* Higher levels of Neanderthal ancestry in East Asians than in Europeans. *Genetics* **194**, 199–209 (2013).
28. Abi-Rached, L. *et al.* The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* **334**, 89–94 (2011).
29. Waddell, P. J. & Tan, X. New *g<sup>2</sup>AIC*, *g<sup>2</sup>AICc*, *g<sup>2</sup>BIC*, and power divergence fit statistics expose mating between modern humans, Neanderthals and other archaics. Preprint at <http://arxiv.org/abs/1212.6820> (2012).
30. Wegmann, D., Leuenberger, C., Neuenschwander, S. & Excoffier, L. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* **11**, 116 (2010).
31. Kumar, R. A. *et al.* Recurrent 16p11.2 microdeletions in autism. *Hum. Mol. Genet.* **17**, 628–638 (2008).
32. Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
33. Fietz, S. A. & Huttner, W. B. Cortical progenitor expansion, self-renewal and neurogenesis—a polarized perspective. *Curr. Opin. Neurobiol.* **21**, 23–35 (2011).
34. Kokovay, E. *et al.* VCAM1 is essential to maintain the structure of the SVZ niche and acts as an environmental sensor to regulate SVZ lineage progression. *Cell Stem Cell* **11**, 220–230 (2012).
35. Wang, C., Liang, C. C., Bian, Z. C., Zhu, Y. & Guan, J. L. FIP200 is required for maintenance and differentiation of postnatal neural stem cells. *Nature Neurosci.* **16**, 532–542 (2013).
36. Rios, D. *et al.* A database and API for variation, dense genotyping and resequencing data. *BMC Bioinformatics* **11**, 238 (2010).
37. Hublin, J. J. Out of Africa: modern human origins special feature: the origin of Neandertals. *Proc. Natl Acad. Sci. USA* **106**, 16022–16027 (2009).
38. Sankararaman, S., Patterson, N., Li, H., Pääbo, S. & Reich, D. The date of interbreeding between Neandertals and modern humans. *PLoS Genet.* **8**, e1002947 (2012).
39. Yang, M. A., Malaspina, A. S., Durand, E. Y. & Slatkin, M. Ancient structure in Africa unlikely to explain Neanderthal and non-African genetic similarity. *Mol. Biol. Evol.* **29**, 2987–2995 (2012).
40. Eriksson, A. & Manica, A. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc. Natl Acad. Sci. USA* **109**, 13956–13960 (2012).
41. Krause, J. *et al.* The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894–897 (2010).
42. Gabunia, L. *et al.* Dmanisi and dispersal. *Evol. Anthropol.* **10**, 158–170 (2001).
43. Asfaw, B. *et al.* Remains of *Homo erectus* from Bouri, Middle Awash, Ethiopia. *Nature* **416**, 317–320 (2002).
44. Kircher, M., Stenzel, U. & Kelso, J. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome Biol.* **10**, R83 (2009).
45. Langergraber, K. E. *et al.* Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl Acad. Sci. USA* **109**, 15716–15721 (2012).
46. Scally, A. & Durbin, R. Revising the human mutation rate: implications for understanding human evolution. *Nature Rev. Genet.* **13**, 745–753 (2012).

**Supplementary Information** is available in the online version of the paper.

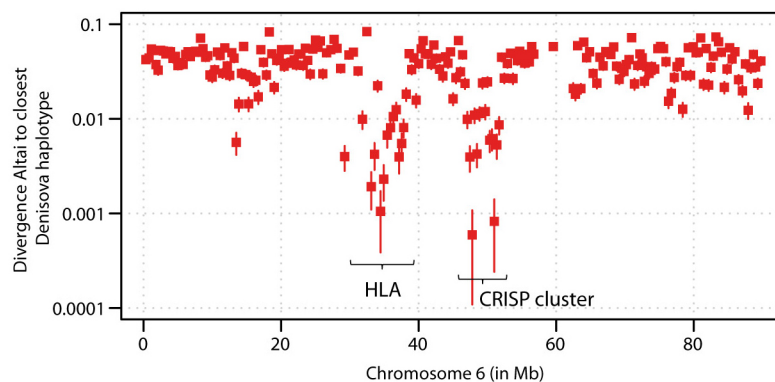
**Acknowledgements** We thank M. Hammer, C. Winkler and W. Klitz for sharing DNA samples; W. Huttner and his group, B. Peter, J. G. Schraiber and M. A. Yang for helpful discussions; and A. Lewis and R. Qiu for technical assistance. N.P. and D.R. are grateful for the chance to discuss these results with Peter Waddell who independently found evidence of a deeply diverged hominin admixing into the Denisova genome. D.R. and E.E.E. are Howard Hughes Medical Institute Investigators. D.R. and N.P. were supported by NSF grant number 1032255 and NIH grant GM100233; E.E.E. by NIH grant HG002385; J.S. by grant HG006283 from the National Genome Research Institute (NHGRI); S.S. by a post-doctoral fellowship from the Harvard University Science of the Human Past Program; F.J. and M.S. in part by a grant from the NIH (R01-GM40282); P.H.S. by an HHMI International Student Fellowship. We thank the team at the NIH Intramural Sequencing Center and Alice Young in particular, for generating some of the sequence reported here. This research was supported in part by the Paul G. Allen Family Foundation. Major funding support came from the Presidential Innovation Fund of the Max Planck Society.

**Author Contributions** S.Saw., A.H. and Q.F. performed the experiments; K.P., F.R., N.P., F.J., S.San., S.Saw., A.H., G.R., P.H.S., C.d.F., M.D., Q.F., M.Ki., M.Ku., M.L., M.M., M.O., M.Si., C.T., H.L., S.M., A.T., P.M., J.P., J.C.M., S.H.V., R.E.G., I.H., P.L.F.J., J.O.K., J.S., E.E.E., E.S.L., T.E.B., M.Si., D.R., J.K., and S.P. analysed genetic data; L.V.G., V.B.D., M.V.S., A.P.D. and B.V. analysed archaeological and anthropological data; H.B. and H.C. provided samples and reagents; K.P., J.K. and S.P. wrote and edited the manuscript with input from all authors.

**Author Information** All sequence data have been submitted to the European Nucleotide Archive (ENA) and are available under the following accessions: Altai Neanderthal: ERP002097, Mezmaiskaya Neanderthal: ERP002447. The data from the 25 present-day human genomes and 13 experimentally phased present-day genomes are available as a public dataset from <http://aws.amazon.com/datasets/> and from <http://cdna.eva.mpg.de/neandertal/altai/>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.S. (slatkin@berkeley.edu), D.R. (reich@genetics.med.harvard.edu) or S.P. (paabo@eva.mpg.de).



**Extended Data Figure 1 | Heterozygosity estimates for the Altai Neanderthal individual, the Denisovan individual, non-Africans and Africans.** The bars for the latter two give the range of heterozygosity observed among 15 non-African and 10 African individuals, respectively (Supplementary Information section 9).



**Extended Data Figure 2 | Neanderthal-introgressed loci in Denisova.**

Divergence of the Altai Neanderthal to the most closely related Denisovan haplotype in windows of at least 200 kb on chromosome 6. Divergence is given

as percentage of human–chimpanzee divergence and bars represent  $\pm 1$  standard error.



Extended Data Table 1 | Neanderthal ancestry estimate

	Other Neanderthal = Mezmaiskaya				Other Neanderthal = Vindija			
	Panel A		Panel B		Panel A		Panel B	
	$\hat{a}$	Std. Err.	$\hat{a}$	Std. Err.	$\hat{a}$	Std. Err.	$\hat{a}$	Std. Err.
French	0.020	0.003	0.019	0.003	0.016	0.002	0.017	0.002
Sardinian	0.019	0.002	0.017	0.003	0.018	0.002	0.018	0.002
Han	0.022	0.003	0.018	0.003	0.023	0.002	0.019	0.002
Dai	0.019	0.003	0.016	0.003	0.019	0.002	0.016	0.002
Karitiana	0.020	0.003	0.019	0.003	0.018	0.002	0.019	0.002
Mixe	-	-	0.018	0.003	-	-	0.017	0.002

Note: we estimate ancestry using the equation  $\hat{z} = \frac{f_4(\text{Denisova, Altai, Africa, X})}{f_4(\text{Denisova, Altai, Africa, Other Neanderthal})}$

**Extended Data Table 2 | Selected *D*-statistics supporting inferences about gene flows**

Statistic	D	Z	Interpretation
D(French, Dinka; Altai, Chimp)	5.4%	9.2	SI 14: Neanderthals share more derived alleles with non-Africans than with Africans.
D(Han, Dinka; Altai, Chimp)	7.3%	11.4	
D(Han, Papuan; Denisova, Chimp)	-7.0%	-9.5	SI 14: Denisovans share more derived alleles with Oceanian populations than with other non Africans.
D(Han, Australian; Denisova, Chimp)	-7.7%	-10.7	
D(Altai, Mezmaiskaya; French, Dinka)	-16.4%	-5.8	SI 14: The archaic material in non-Africans falls within late Neanderthal variation: Non-Africans share more alleles with some Neanderthals (Mezmaiskaya/Vindija) than others (Altai).
D(Altai, Vindija; French, Dinka)	-7.0%	-4.3	
D(Altai, Mezmaiskaya; Denisova, Chimp)	13.2%	5.9	SI 15: Gene flow between Altai related Neanderthals and Denisovans (Denisovans share more derived alleles with Altai than with Mezmaiskaya)
D(Altai, Vindija; Denisova, Chimp)	7.9%	5.6	
D(Altai, Denisova; 12 Africans, Chimp)	7.0%	11.6	SI 16: Unknown archaic gene flow into Denisova: Africans share more derived alleles with Altai than with Denisova, a signal that strengthens for fixed derived alleles
D(Altai, Denisova; 12 Africans Fixed, Chimp)	13.4%	10.0	

**Extended Data Table 3 | Lineage-specific segmental duplications along each of the terminal branches and genes encompassed**

Locus	Length	Lineage	Genes	Genotypes			
				Modern Humans (median)	Denisova	Altai	Mezmai skaya
chr12:122079832-122087495	7663	Altai-Neanderthal	ORAI1	2	2	4	3
chr12:132295389-132391442	96053	Altai-Neanderthal	MMP17,ULK1	2	2	4	2
chr19:9284044-9291195	7151	Altai-Neanderthal		2	2	4	4
chr20:281880-290717	8837	Altai-Neanderthal		2	2	10	9
chr3:12639069-12641393	2324	Altai-Neanderthal	RAF1	2	2	7	3
chr6:95473793-95532866	59073	Altai-Neanderthal		2	2	3	2
chr11:39901956-39909545	7589	Denisova		2	4	2	2
chr1:161272681-161274838	2157	Denisova	MPZ	2	4	2	2
chr12:49894191-49897733	3542	Denisova	SPATS2	2	4	2	2
chr19:55302094-55315197	13103	Denisova	KIR3DP1,KIR2DL4	2	4	2	2
chr2:48781187-48787915	6728	Denisova		2	3	2	2
chr4:68542692-68577288	34596	Denisova	UBA6,LOC550112	2	3	2	2
chr4:68579206-68581585	2379	Denisova	LOC550112	2	3	2	2
chr7:140872574-140879065	6491	Denisova	LOC100131199	2	6	2	2
chr1:108924526-108990191	65665	Modern Human		4	2	2	2
chr16:30200098-30206185	6087	Modern Human	CORO1A,LOC606724,BOLA2	6	2	2	2
chr2:87417089-87420544	3455	Modern Human		4	2	2	2



# A molecular marker of artemisinin-resistant *Plasmodium falciparum* malaria

Frédéric Arie<sup>1,2†</sup>, Benoit Witkowski<sup>3</sup>, Chanaki Amaratunga<sup>4</sup>, Johann Beghain<sup>1,2†</sup>, Anne-Claire Langlois<sup>1,2</sup>, Nimol Khim<sup>3</sup>, Saorin Kim<sup>3</sup>, Valentine Duru<sup>3</sup>, Christiane Bouchier<sup>5</sup>, Laurence Ma<sup>5</sup>, Pharath Lim<sup>3,4,6</sup>, Rithea Leang<sup>6</sup>, Socheat Duong<sup>6</sup>, Sokunthea Sreng<sup>6</sup>, Seila Suon<sup>6</sup>, Char Meng Chuor<sup>6</sup>, Denis Mey Bout<sup>7</sup>, Sandie Ménard<sup>8†</sup>, William O. Rogers<sup>9</sup>, Blaise Genton<sup>10</sup>, Thierry Fandeur<sup>1,3</sup>, Olivo Miotto<sup>11,12,13</sup>, Pascal Ringwald<sup>14</sup>, Jacques Le Bras<sup>15</sup>, Antoine Berry<sup>8†</sup>, Jean-Christophe Barale<sup>1,2†</sup>, Rick M. Fairhurst<sup>4\*</sup>, Françoise Benoit-Vical<sup>16,17\*</sup>, Odile Mercereau-Puijalon<sup>1,2\*</sup> & Didier Ménard<sup>3\*</sup>

***Plasmodium falciparum* resistance to artemisinin derivatives in southeast Asia threatens malaria control and elimination activities worldwide. To monitor the spread of artemisinin resistance, a molecular marker is urgently needed. Here, using whole-genome sequencing of an artemisinin-resistant parasite line from Africa and clinical parasite isolates from Cambodia, we associate mutations in the PF3D7\_1343700 kelch propeller domain ('K13-propeller') with artemisinin resistance *in vitro* and *in vivo*. Mutant K13-propeller alleles cluster in Cambodian provinces where resistance is prevalent, and the increasing frequency of a dominant mutant K13-propeller allele correlates with the recent spread of resistance in western Cambodia. Strong correlations between the presence of a mutant allele, *in vitro* parasite survival rates and *in vivo* parasite clearance rates indicate that K13-propeller mutations are important determinants of artemisinin resistance. K13-propeller polymorphism constitutes a useful molecular marker for large-scale surveillance efforts to contain artemisinin resistance in the Greater Mekong Subregion and prevent its global spread.**

The emergence of *Plasmodium falciparum* resistance to artemisinin derivatives (ART) in Cambodia threatens the world's malaria control and elimination efforts<sup>1,2</sup>. The risk of ART-resistant parasites spreading from western Cambodia to the Greater Mekong Subregion and to Africa, as happened previously with chloroquine- and sulphadoxine/pyrimethamine-resistant parasites<sup>3–5</sup>, is extremely worrisome. Clinical ART resistance is defined as a reduced parasite clearance rate<sup>1,6–10</sup>, expressed as an increased parasite clearance half-life<sup>11,12</sup>, or a persistence of microscopically detectable parasites on the third day of artemisinin-based combination therapy (ACT)<sup>2</sup>. The half-life parameter correlates strongly with results from the *in vitro* ring-stage survival assay (RSA<sub>0–3h</sub>) and results from the *ex vivo* RSA<sup>13</sup>, which measure the survival rate of young ring-stage parasites to a pharmacologically relevant exposure (700 nM for 6 h) to dihydroartemisinin (DHA)—the major metabolite of all ARTs. However, the present lack of a molecular marker hampers focused containment of ART-resistant parasites in areas where they have been documented and hinders rapid detection of these parasites elsewhere, where ACTs remain the most affordable, effective antimalarials. To detect and monitor the spread of ART resistance, a molecular marker for widespread use is needed.

Recent genome-wide analyses of *P. falciparum* isolates have provided evidence of recent positive selection in geographic areas of ART resistance<sup>9,14–16</sup>. Whereas parasite heritability of the clinical phenotype

is above 50%, no reliable molecular marker has yet been identified. One possible explanation is that the parasite clearance half-life is not only determined by the intrinsic susceptibility of a parasite isolate to ART, but also by its developmental stage at the time of ART treatment and host-related parameters such as pharmacokinetics and immunity<sup>17</sup>. This issue was recently highlighted in patients presenting discordant data between parasite clearance half-life *in vivo* and RSA<sub>0–3h</sub> survival rate *in vitro*<sup>13</sup>. Moreover, genome-wide association studies (GWAS) are confounded by uncertainties about parasite population structure. Recent evidence for several highly differentiated subpopulations of ART-resistant parasites in western Cambodia<sup>15</sup> suggests that distinct emergence events might be occurring. An alternative strategy to discover a molecular marker is to analyse mutations acquired specifically by laboratory-adapted parasite clones selected to survive high doses of ART *in vitro*, and use this information to guide analysis of polymorphism in clinical parasite isolates from areas where ART resistance is well documented at both temporal and geographical levels. Here we used this strategy to explore the molecular signatures of clinical ART resistance in Cambodia, where this phenotype was first reported<sup>1,8</sup>.

## A candidate molecular marker of ART resistance

The ART-resistant F32-ART5 parasite line was selected by culturing the ART-sensitive F32-Tanzania clone under a dose-escalating, 125-cycle

<sup>1</sup>Institut Pasteur, Parasite Molecular Immunology Unit, 75724 Paris Cedex 15, France. <sup>2</sup>Centre National de la Recherche Scientifique, Unité de Recherche Associée 2581, 75724 Paris Cedex 15, France.

<sup>3</sup>Institut Pasteur du Cambodge, Malaria Molecular Epidemiology Unit, Phnom Penh, Cambodia. <sup>4</sup>Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>5</sup>Institut Pasteur, Plate-forme Génomique, Département Génomique et Génétique, 75724 Paris Cedex 15, France. <sup>6</sup>National Center for Parasitology, Entomology and Malaria Control, Phnom Penh, Cambodia. <sup>7</sup>SSA WHO, Drug Monitoring in Cambodia, National Center for Parasitology, Entomology and Malaria Control, Phnom Penh, Cambodia. <sup>8</sup>Service de Parasitologie et Mycologie, Centre Hospitalier Universitaire de Toulouse, 31059 Toulouse Cedex 9, France. <sup>9</sup>Naval Medical Research Unit #2 Detachment, Phnom Penh, Cambodia. <sup>10</sup>Swiss Tropical and Public Health Institute, 4051 Basel, Switzerland. <sup>11</sup>MRC Centre for Genomics and Global Health, University of Oxford, Oxford OX3 7BN, UK. <sup>12</sup>Mahidol-Oxford Tropical Medicine Research Unit, Mahidol University, Bangkok 10400, Thailand. <sup>13</sup>Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. <sup>14</sup>Global Malaria Program, World Health Organization, 1211 Geneva, Switzerland. <sup>15</sup>Centre National de Référence du Paludisme, CHU Bichat-Claude Bernard, APHP, PRES Sorbonne Paris Cité, 75018 Paris, France. <sup>16</sup>Centre National de la Recherche Scientifique, Laboratoire de Chimie de Coordination UPR8241, 31077 Toulouse Cedex 4, France. <sup>17</sup>Université de Toulouse, UPS, Institut National Polytechnique de Toulouse, 31077 Toulouse Cedex 4, France. <sup>†</sup>Present addresses: Institut Pasteur, Genetics and Genomics of Insect Vectors Unit, 75724 Paris Cedex 15, France (F.A.); Institut Pasteur, Functional Genetics of Infectious Diseases Unit, 75724 Paris Cedex 15, France (J.B.); Centre de Physiopathologie de Toulouse-Purpan, Institut National de la Santé et de la Recherche Médicale UMR1043, Centre National de la Recherche Scientifique UMR5282, Université Toulouse III, 31024 Toulouse Cedex 3, France (S.M. & A.B.); Institut Pasteur, Unité de Biologie et Génétique du Paludisme, Team Malaria Targets and Drug Development, 75724 Paris Cedex 15, France (J.-C.B.).

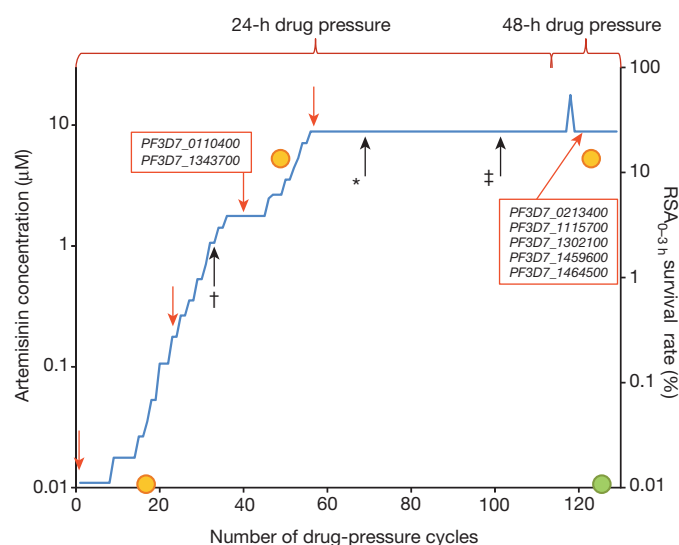
\*These authors contributed equally to this work.

regimen of artemisinin for 5 years<sup>18</sup>. Whole-genome sequences were obtained for both F32-ART5 and F32-TEM (its sibling clone cultured without artemisinin) at 460× and 500× average nucleotide coverage, respectively. Compared to F32-TEM, no deleted genes were identified in F32-ART5. The exomes of F32-ART5 and F32-TEM were compared after excluding (1) genes from highly variable, multi-gene families (*var*, *rifin* and *stevor*), (2) positions with coverage lower than 25% of the mean coverage of the parasite line, (3) single-nucleotide polymorphisms (SNPs) found to be mixed in F32-ART5, given that acquired ART-resistance mutation(s) could be expected to be fixed in the sample after 5 years of continuous pressure, (4) SNPs shared between F32-ART5 and the ART-sensitive 3D7 parasite strain and (5) synonymous SNPs (Extended Data Fig. 1).

This analysis identified eight mutations in seven genes that were subsequently confirmed by Sanger sequencing of PCR products (Extended Data Table 1). Each gene harbours one mutant codon in F32-ART5 compared to F32-TEM, F32-Tanzania or 3D7 (Extended Data Table 2). Information on the expression of the genes and the biological function of the proteins are listed in Extended Data Table 3. Only one of these genes, cysteine protease falcipain 2a (*PF3D7\_1115700*), has previously been associated with *in vitro* responses to ART<sup>19</sup>. To determine when each mutation arose in the F32-ART5 lineage, we analysed the whole-genome sequences of parasites at various drug-pressure cycles (Fig. 1). This analysis showed that the *PF3D7\_0110400* D56V and *PF3D7\_1343700* M476I mutations were acquired first, during the steep increase of ART resistance, and remained stable thereafter. Importantly, the appearance of these two mutations is associated with an increase in the RSA<sub>0–3h</sub> survival rate, from less than 0.01% to 12.8%. Subsequent PCR analysis of the *PF3D7\_1343700* locus detected the M476I mutation after 30 drug-pressure cycles, consistent with the sharp increase in RSA<sub>0–3h</sub> survival rate observed thereafter. The other SNPs appeared stepwise at later stages of selection: *PF3D7\_0213400* (68 cycles); *PF3D7\_1115700* (98 cycles); *PF3D7\_1302100*, *PF3D7\_1459600* and *PF3D7\_1464500* (120 cycles) (Extended Data Table 2). These data indicate that the *PF3D7\_1343700* M476I mutation increased the resistance of F32-Tanzania to DHA in the RSA<sub>0–3h</sub>.

To explore whether these mutations are associated with ART resistance in Cambodia, we investigated sequence polymorphism in all seven genes by mining whole-genome or Sanger sequences for 49 culture-adapted parasite isolates collected in 2010–2011 (see Methods). We chose these isolates based on their differential RSA<sub>0–3h</sub> survival rates (Supplementary Table 1) and their sequences were compared to those of control parasite lines 3D7, 89F5<sup>20</sup> and K1992 (see Methods). Three genes (*PF3D7\_0110400*, *PF3D7\_0213400* and *PF3D7\_1302100*) encode a wild-type sequence for all parasite isolates. The other four genes show intra-population diversity, with previously reported or novel SNPs (Supplementary Table 1). *PF3D7\_1115700* has 11 SNPs that are not associated with RSA<sub>0–3h</sub> survival rates ( $P = 0.06$ , Kruskal–Wallis test). *PF3D7\_1459600* has 6 SNPs that are not associated with survival rates ( $P = 0.65$ ). *PF3D7\_1464500* has 12 SNPs previously reported in older isolates from southeast Asia, including the ART-susceptible Dd2 line<sup>21</sup>, probably reflecting a geographic signature. These SNPs also show no significant association with survival rates ( $P = 0.42$ ). Therefore, these six genes were not studied further.

In contrast, *PF3D7\_1343700* polymorphism shows a significant association with RSA<sub>0–3h</sub> survival rates (Fig. 2). Indeed, RSA<sub>0–3h</sub> survival rates differ substantially between parasite isolates with wild-type (median 0.17%, range 0.06–0.51%,  $n = 16$ ) or mutant (18.8%, 3.8–58%,  $n = 33$ ) K13-propeller alleles ( $P < 10^{-4}$ , Mann–Whitney  $U$  test) (Supplementary Table 1). Four mutant alleles are observed, each harbouring a single non-synonymous SNP within a kelch repeat of the C-terminal K13-propeller domain, namely Y493H, R539T, I543T and C580Y located within repeats no. 2, 3, 3 and 4, respectively. Both the K1992 and the ART-susceptible 89F5 lines carry a wild-type K13-propeller. There are no associations between polymorphisms in the K13-propeller and those in the other candidate genes (Supplementary Table 1). Based on these

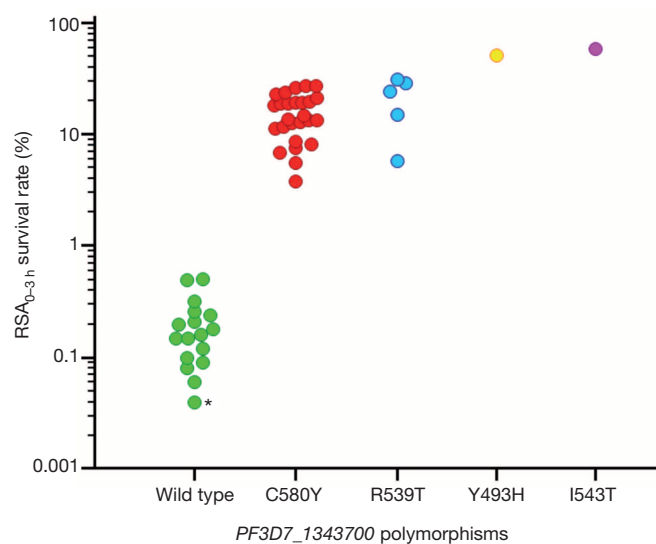


**Figure 1 | Temporal acquisition of mutations in F32-ART5.** F32-Tanzania parasites exposed to increasing artemisinin concentrations for 120 consecutive cycles<sup>18</sup> were analysed by whole-genome sequencing at five time-points (red arrows). Loci mutated after a given number of drug-pressure cycles are shown (red boxes). The earliest time-points where three mutations were detected by PCR (black arrows) are indicated by † for *PF3D7\_1343700*, \* for *PF3D7\_0213400* and ‡ for *PF3D7\_1115700*. Orange and green circles indicate RSA<sub>0–3h</sub> survival rates for F32-ART5 and F32-TEM parasites, respectively (mean of 3 experiments each).

observations and the acquisition of M476I in kelch repeat no. 2 by F32-ART5, we investigated whether K13-propeller polymorphism is a molecular signature of ART resistance in Cambodia.

### Emergence and spread of K13-propeller mutants in Cambodia

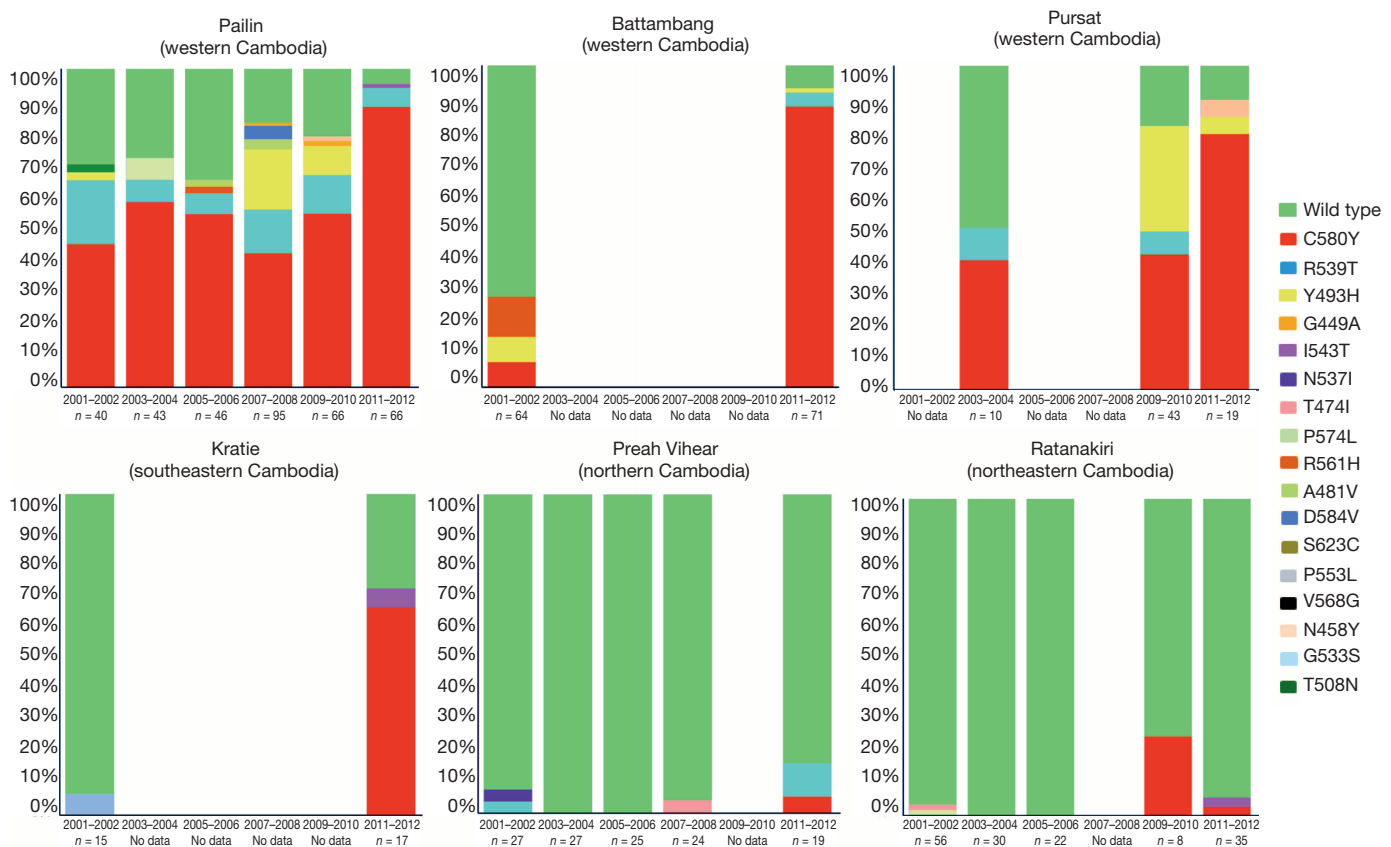
Over the last decade, the prevalence of ART resistance has steadily increased in the western provinces of Cambodia, but not elsewhere in



**Figure 2 | Survival rates of Cambodian parasite isolates in the RSA<sub>0–3h</sub> stratified by K13-propeller allele.** Genotypes were obtained by mining whole-genome sequence data ( $n = 21$ ) or sequencing PCR products ( $n = 28$ ). Mutant parasites have significantly higher RSA<sub>0–3h</sub> survival rates than wild-type parasites: wild type ( $n = 17$ , median 0.16%, IQR 0.09–0.24, range 0.04–0.51); C580Y ( $n = 26$ , median 14.1%, IQR 11.3–19.6, range 3.8–27.3,  $P < 10^{-6}$  for wild type versus C580Y, Mann–Whitney  $U$  test); R539T ( $n = 5$ , median 24.2%, IQR 12.6–29.5, range 5.8–31.3,  $P < 10^{-3}$  for wild type versus R539T); Y493H (51.4%); and I543T (58.0%). The RSA<sub>0–3h</sub> survival rate (0.04%) of control 3D7 parasites is indicated by an asterisk.

the country<sup>2</sup>. To test whether the spatiotemporal distribution of K13-propeller mutations correlates with that of ART resistance, we sequenced the K13-propeller of archived parasite isolates from Cambodian patients with malaria in 2001–2012 (Extended Data Table 4). Data from six provinces were compared ( $n = 886$ ): Pailin, Battambang and Pursat in the west where ART resistance is established<sup>1,6,8,22</sup>, Kratie in the south-east where ART resistance has increased in recent years<sup>2</sup>, and Preah Vihear in the north and Ratanakiri in the northeast where there was virtually no evidence of ART resistance during this time period<sup>2</sup>. This analysis reveals overall 17 mutant alleles, including three high-frequency ( $> 5\%$ ) alleles (C580Y, R539T and Y493H). The frequency of wild-type sequence decreased significantly over time in all three western provinces, but not in Preah Vihear or Ratanakiri. The frequency of the C580Y allele increased significantly from 2001–2002 to 2011–2012 in Pailin and Battambang, indicating its rapid invasion of the population and near fixation in these areas (Fig. 3).

To further investigate the geographic diversity of K13-propeller polymorphism in Cambodia, we extended our sequence analysis to include data from four additional provinces ( $n = 55$ , Kampong Som, Kampot, Mondulkiri and Oddar Meanchey) in 2011–2012 (Extended Data Table 4). Although a large number of mutations are observed (Supplementary Fig. 1 and Extended Data Table 5), the C580Y allele accounts for 85% (189/222) of all mutant alleles observed in 2011–2012 (Extended Data Fig. 2). This mapping outlines the elevated frequency (74%, 222/300) of parasites harbouring a single non-synonymous mutation in the K13-propeller and the geographic disparity of their distribution. Importantly, the frequency distribution of mutant alleles over the various provinces matches that of day 3 positivity in patients treated for malaria with an ACT (Spearman's  $r = 0.99$ , 95% confidence interval 0.96–0.99,  $P < 0.0001$ ), considered a suggestive sign of clinical ART resistance (Extended Data Fig. 3).



**Figure 3 | Frequency of K13-propeller alleles in 886 parasite isolates in six Cambodian provinces in 2001–2012.** Genotypes were obtained by sequencing PCR products from archived blood samples. All mutant alleles carry a single non-synonymous SNP (colour-coded, same colour codes as in

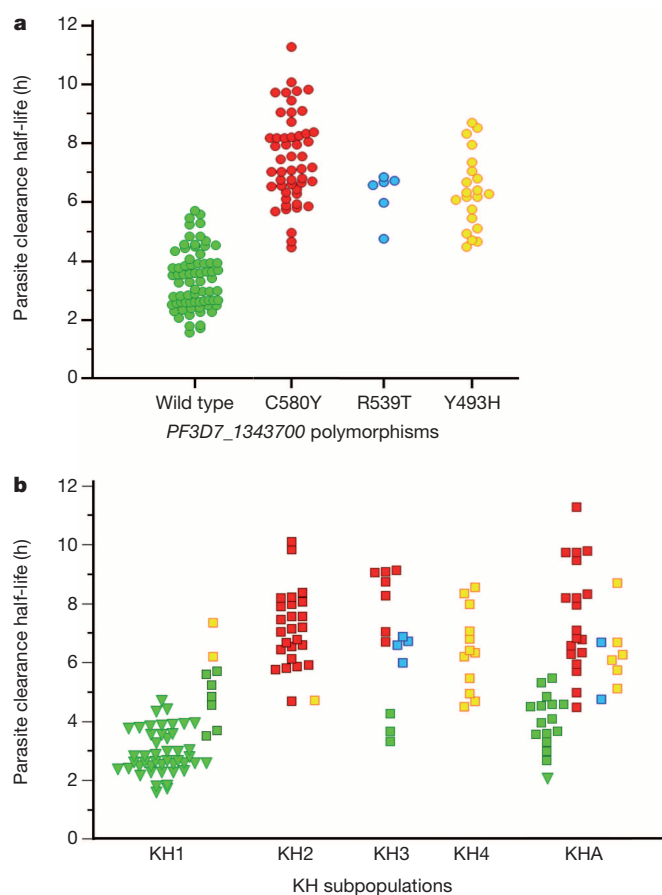
### K13-propeller polymorphisms and clinical ART resistance

To confirm that K13-propeller polymorphism is a molecular marker of clinical ART resistance, we first identified 163 patients from Pursat and Ratanakiri in whom we measured parasite clearance half-lives (range 1.58–11.53 h)<sup>6</sup> in 2009–2010 and for which parasites were previously assigned to a KH subpopulation (KH1, KH2, KH3, KH4 or KHA) on the basis of ancestry analysis of whole-genome sequence data<sup>15</sup>. Thirteen patients with mixed genotypes (a wild-type and one or more mutant K13-propeller alleles) were excluded. Of the remaining 150 patients, 72 carried parasites with a wild-type allele and the others carried parasites with only a single non-synonymous SNP in the K13-propeller: C580Y ( $n = 51$ ), R539T ( $n = 6$ ) and Y493H ( $n = 21$ ) (Extended Data Table 6). The parasite clearance half-life in patients with wild-type parasites is significantly shorter (median 3.30 h, interquartile range (IQR) 2.59–3.95) than those with C580Y (7.19 h, 6.47–8.31,  $P < 10^{-6}$ , Mann–Whitney  $U$  test), R539T (6.64 h, 6.00–6.72,  $P < 10^{-4}$ ) or Y493H (6.28 h, 5.37–7.14,  $P < 10^{-6}$ ) parasites (Fig. 4a). Also, the parasite clearance half-life in patients carrying C580Y parasites is significantly longer than those with Y493H parasites ( $P = 0.007$ , Mann–Whitney  $U$  test). These data indicate that C580Y, R539T and Y493H identify slow-clearing parasites in malaria patients treated with ART.

Because KH2, KH3, KH4 and KHA parasites have longer half-lives than KH1 parasites<sup>15</sup>, we proposed that allelic variation in the K13-propeller accounts for these differences. Among 150 parasites, 55, 26, 14, 12 and 43 are classified as KH1, KH2, KH3, KH4 and KHA, respectively. Three K13-propeller alleles strongly associate with KH groups: 96% (53/55) of KH1, 96% (25/26) of KH2 and 100% (12/12) of KH4 parasites carry the wild-type, C580Y and Y493H alleles, respectively (Extended Data Table 6). Whereas KH3 parasites ( $n = 14$ ) carry the wild-type, C580Y and R539T alleles, R539T is not observed in KH1, KH2 or KH4

Fig. 2 for wild type, C580Y, R539T, Y493H and I543T). Significant reductions (Fisher's exact test) in wild-type allele frequencies were observed in Pailin, Battambang, Pursat and Kratie over time (see Methods).





**Figure 4 | Parasite clearance half-lives.** **a**, Correlation of parasite clearance half-lives and K13-propeller alleles for parasite isolates in Pursat and Ratanakiri in 2009–2010. Wild-type parasites have shorter half-lives (median 3.30 h, IQR 2.59–3.95,  $n = 72$ ) than C580Y (7.19 h, 6.47–8.31,  $n = 51$ ,  $P < 10^{-6}$ , Mann–Whitney  $U$  test), R539T (6.64 h, 6.00–6.72,  $n = 6$ ,  $P < 10^{-6}$ ) or Y493H (6.28 h, 5.37–7.14,  $n = 21$ ,  $P < 10^{-6}$ ) parasites. The half-life of C580Y parasites is significantly longer than that of Y493H parasites ( $P = 0.007$ ). **b**, Correlation of parasite clearance half-lives, KH subpopulations<sup>15</sup> and K13-propeller alleles for the same 150 parasite isolates. Half-lives are shown for Pursat (squares) and Ratanakiri (triangles) parasites, stratified by KH group and K13-propeller allele (colour-coded as in **a**). Median half-lives stratified by K13-propeller allele are KH1: wild type (2.88) and Y493H (6.77); KH2: C580Y (7.13) and Y493H (4.71); KH3: wild type (3.65), C580Y (8.73) and R539T (6.65); KH4: Y493H (6.37); and KHA: wild type (4.01), C580Y (7.09), Y493H (6.18) and R539T (5.73).

parasites. As expected, KHA parasites have a mixed allele composition. Importantly, K13-propeller mutations more accurately identify slow-clearing parasites than KH group (Fig. 4b), demonstrating that the association of K13-propeller polymorphism with clinical ART resistance in Cambodia is partially independent of the genetic background of KH subpopulations. Within the KH1 group ( $n = 55$ ), the parasite clearance half-life in patients with wild-type parasites is significantly shorter ( $n = 53$ , median 2.88 h, IQR 2.52–3.79) than those with Y493H parasites ( $n = 2$ , median 6.77 h,  $P = 0.02$ , Mann–Whitney  $U$  test). Within the KH3 subpopulation ( $n = 14$ ), the half-life in patients with wild-type parasites is shorter ( $n = 3$ , median 3.65 h) than those with C580Y ( $n = 7$ , median 8.73 h, IQR 7.35–9.06,  $P = 0.02$ ) or R539T ( $n = 4$ , 6.65 h, 6.29–6.80,  $P = 0.03$ ) parasites.

## Discussion

The F32-ART5 lineage acquired a K13-propeller mutation as it developed ART resistance, as indicated by its ability to survive a pharmacologically relevant exposure to DHA in the RSA<sub>0–3h</sub>. Genes putatively associated with ART resistance (*Pfcr*<sup>23,24</sup>, *Pfctcp*<sup>25,26</sup>, *Pfmdr1*<sup>8,27,28</sup>, *Pfmrp1*<sup>27–29</sup> and

ABC transporters<sup>30</sup>) or encoding putative targets of ART (*PfATPase6*<sup>31,32</sup> and *Pfubctb*—the orthologue of *Plasmodium chabaudi* ubp1<sup>33,34</sup>) were not mutated during the 5-year selection of F32-ART5, and *Pfmdr1* amplification was not observed<sup>35–40</sup>. In addition, all candidate ART-resistance genes recently identified using population genetics approaches<sup>14,40,41</sup> remained unaltered in F32-ART5, except for *PF3D7\_1343700* and *PF3D7\_1459600* located in the linkage-disequilibrium windows identified in ref. 16. These findings led us to identify another 17 single K13-propeller mutations in naturally circulating parasites in Cambodia. Several of these mutations associate strongly with the spatiotemporal distribution of ART resistance in Cambodia, increased parasite survival rates in response to DHA *in vitro*, and long parasite clearance half-lives in response to ART treatment *in vivo*. None of the six other genes mutated in F32-ART5 associate with RSA<sub>0–3h</sub> survival rates in parasite isolates from Cambodia.

K13-propeller polymorphism fulfils the definition of a molecular marker of ART resistance for several reasons: (1) there has been a progressive loss of wild-type parasites in western Cambodia during the decade of emerging ART resistance in this region; (2) mutant parasites cluster in Cambodian provinces where ART resistance is well established and are less prevalent where ART resistance is uncommon; (3) *PF3D7\_1343700* is located 5.9 kilobases upstream of the 35-kb locus identified in ref. 14 as being under recent positive selection, and within the region of top-ranked signatures of selection outlined in ref. 16; (4) multiple mutations, all non-synonymous, are present in the K13-propeller, reflecting positive selection rather than a hitchhiking effect or genetic drift; (5) mutations occur in a domain that is highly conserved in *P. falciparum*, with only one non-synonymous SNP being documented in a single parasite isolate from Africa<sup>42</sup>; (6) all polymorphisms we observe in Cambodia are novel and all but one (V568G) occur at positions strictly conserved between *Plasmodium* species (Supplementary Fig. 1 and Supplementary Fig. 2), suggesting strong structural and functional constraints on the protein; (7) the three most-prevalent K13-propeller mutations correlate strongly with RSA<sub>0–3h</sub> survival rates *in vitro* and parasite clearance half-lives *in vivo* at the level of individual parasite isolates and malaria patients, respectively; and (8) the frequency of mutant alleles correlates strongly with the prevalence of day 3 positivity after ACT treatment at the level of human populations in Cambodia.

On the basis of homology with other kelch propeller domains, we anticipate that the observed K13-propeller mutations destabilize the domain scaffold and alter its function. The carboxy-terminal portion of *PF3D7\_1343700* encodes six kelch motifs, which are found in a large number of proteins with diverse cellular functions<sup>43,44</sup>. Given that the toxicity of ART derivatives depends principally on their pro-oxidant activity, the reported role of some kelch-containing proteins in regulating cytoprotective and protein degradation responses to external stress is particularly intriguing. The K13-propeller shows homology with human KLHL12 and KLHL2, involved in ubiquitin-based protein degradation, and KEAP1, involved in cell adaptation to oxidative stress (Extended Data Fig. 4). More work is needed to delineate the normal function of K13 and the effect of various mutations. Allele exchange studies in mutant and wild-type parasites may help to define the contribution of K13-propeller polymorphisms on different genetic backgrounds to the RSA<sub>0–3h</sub> survival rate. Indeed, it is particularly worrying that as few as two mutations, that is, the K13-propeller M476I and *PF3D7\_0110400*D56V, were sufficient to confer ART resistance to F32-Tanzania, which has a typical African genetic background. Cambodian parasites with mutant K13-propellers display a wide range of RSA<sub>0–3h</sub> survival rates (3.8–58%) and parasite clearance half-lives (4.5–11.5 h). Further studies are therefore required to identify additional genetic determinants of ART resistance, which may reside in the strongly selected regions recently identified<sup>14,16</sup>. In this context, analysing the RSA<sub>0–3h</sub> survival rates as a quantitative trait among parasites harbouring the same K13-propeller mutation could help to identify additional genetic loci involved in ART resistance.

In summary, K13-propeller polymorphism seems to be a useful molecular marker for tracking the emergence and spread of ART-resistant *P. falciparum*.

## METHODS SUMMARY

The ART-resistant F32-ART5 parasite line was selected by culturing the ART-sensitive F32-Tanzania clone under a dose-escalating regimen of artemisinin for 5 years. The F32-TEM line was obtained by culturing F32-Tanzania in parallel without artemisinin exposure. Reference DNA was extracted from *P. falciparum* lines 3D7, 89F5 Palo Alto Uganda and K1992. The ring-stage survival assay (RSA<sub>0-3h</sub>) was performed as described previously<sup>13</sup>. Whole-genome sequencing was performed on F32-Tanzania, F32-TEM, F32-ART5 (4 time points), three reference strains (3D7, 89F5 and K1992) and 21 Cambodian parasite isolates, using an Illumina paired-reads sequencing technology. A set of 1091 clinical *P. falciparum* isolates was collected from patients participating in ACT efficacy studies in 2001–2012. The K13-propeller was amplified using nested PCR. Double-strand sequencing of PCR products was performed by Macrogen. Sequences were analysed with MEGA 5 software version 5.10 to identify specific SNP combinations. Data were analysed with Microsoft Excel and MedCalc version 12. Differences were considered statistically significant when *P* values were less than 0.05. Ethical clearances for parasite isolate collections were obtained from the Cambodian National Ethics Committee for Health Research, the Institutional Review Board of the Naval Medical Research Center, the Technical Review Group of the WHO Regional Office for the Western Pacific, and the Institutional Review Board of the National Institute of Allergy and Infectious Diseases.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 July; accepted 12 November 2013.

Published online 18 December 2013.

- Dondorp, A. M. *et al.* Artemisinin resistance in *Plasmodium falciparum* malaria. *N. Engl. J. Med.* **361**, 455–467 (2009).
- World Health Organization. *Global Report on Antimalarial Drug Efficacy and Drug Resistance: 2000–2010* (World Health Organization, 2010).
- Mita, T. *et al.* Limited geographical origin and global spread of sulfadoxine-resistant *dhps* alleles in *Plasmodium falciparum* populations. *J. Infect. Dis.* **204**, 1980–1988 (2011).
- Roper, C. *et al.* Intercontinental spread of pyrimethamine-resistant malaria. *Science* **305**, 1124 (2004).
- Wootton, J. C. *et al.* Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* **418**, 320–323 (2002).
- Amaratunga, C. *et al.* Artemisinin-resistant *Plasmodium falciparum* in Pursat province, western Cambodia: a parasite clearance rate study. *Lancet Infect. Dis.* **12**, 851–858 (2012).
- Kyaw, M. P. *et al.* Reduced susceptibility of *Plasmodium falciparum* to artesunate in southern Myanmar. *PLoS ONE* **8**, e57689 (2013).
- Noedl, H. *et al.* Evidence of artemisinin-resistant malaria in western Cambodia. *N. Engl. J. Med.* **359**, 2619–2620 (2008).
- Phyo, A. P. *et al.* Emergence of artemisinin-resistant malaria on the western border of Thailand: a longitudinal study. *Lancet* **379**, 1960–1966 (2012).
- Hien, T. T. *et al.* *In vivo* susceptibility of *Plasmodium falciparum* to artesunate in Binh Phuoc Province, Vietnam. *Malar. J.* **11**, 355 (2012).
- Flegg, J. A. *et al.* Standardizing the measurement of parasite clearance in falciparum malaria: the parasite clearance estimator. *Malar. J.* **10**, 339 (2011).
- White, N. J. The parasite clearance curve. *Malar. J.* **10**, 278 (2011).
- Witkowski, B. *et al.* Novel phenotypic assays for the detection of artemisinin-resistant *Plasmodium falciparum* malaria in Cambodia: in-vitro and ex-vivo drug-response studies. *Lancet Infect. Dis.* **13**, 1043–1049 (2013).
- Cheeseman, I. H. *et al.* A major genome region underlying artemisinin resistance in malaria. *Science* **336**, 79–82 (2012).
- Miotto, O. *et al.* Multiple populations of artemisinin-resistant *Plasmodium falciparum* in Cambodia. *Nature Genet.* **45**, 648–655 (2013).
- Takala-Harrison, S. *et al.* Genetic loci associated with delayed clearance of *Plasmodium falciparum* following artemisinin treatment in Southeast Asia. *Proc. Natl Acad. Sci. USA* **110**, 240–245 (2013).
- Lopera-Mesa, T. M. *et al.* *Plasmodium falciparum* clearance rates in response to artesunate in Malian children with malaria: effect of acquired immunity. *J. Infect. Dis.* **207**, 1655–1663 (2013).
- Witkowski, B. *et al.* Increased tolerance to artemisinin in *Plasmodium falciparum* is mediated by a quiescence mechanism. *Antimicrob. Agents Chemother.* **54**, 1872–1877 (2010).
- Klonis, N. *et al.* Artemisinin activity against *Plasmodium falciparum* requires hemoglobin uptake and digestion. *Proc. Natl Acad. Sci. USA* **108**, 11405–11410 (2011).
- Vigan-Womas, I. *et al.* An in vivo and in vitro model of *Plasmodium falciparum* rosetting and autoagglutination mediated by *varO*, a group A *var* gene encoding a frequent serotype. *Infect. Immun.* **76**, 5565–5580 (2008).
- Cui, L. *et al.* Mechanisms of *in vitro* resistance to dihydroartemisinin in *Plasmodium falciparum*. *Mol. Microbiol.* **86**, 111–128 (2012).
- Leang, R. *et al.* Efficacy of dihydroartemisinin-piperaquine for treatment of uncomplicated *Plasmodium falciparum* and *Plasmodium vivax* in Cambodia, 2008 to 2010. *Antimicrob. Agents Chemother.* **57**, 818–826 (2013).
- Sidhu, A. B. *et al.* Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by *pfcr* mutations. *Science* **298**, 210–213 (2002).
- Valderramos, S. G. *et al.* Identification of a mutant PfCRT-mediated chloroquine tolerance phenotype in *Plasmodium falciparum*. *PLoS Pathog.* **6**, e1000887 (2010).
- Bhisutthibhan, J. *et al.* The *Plasmodium falciparum* translationally controlled tumor protein homolog and its reaction with the antimalarial drug artemisinin. *J. Biol. Chem.* **273**, 16192–16198 (1998).
- Eichhorn, T. *et al.* Molecular interaction of artemisinin with translationally controlled tumor protein (TCTP) of *Plasmodium falciparum*. *Biochem. Pharmacol.* **85**, 38–45 (2013).
- Sanchez, C. P. *et al.* Polymorphisms within PfMDR1 alter the substrate specificity for anti-malarial drugs in *Plasmodium falciparum*. *Mol. Microbiol.* **70**, 786–798 (2008).
- Veiga, M. I. *et al.* Novel polymorphisms in *Plasmodium falciparum* ABC transporter genes are associated with major ACT antimalarial drug resistance. *PLoS ONE* **6**, e20212 (2011).
- Raj, D. K. *et al.* Disruption of a *Plasmodium falciparum* multidrug resistance-associated protein (PfMRP) alters its fitness and transport of antimalarial drugs and glutathione. *J. Biol. Chem.* **284**, 7687–7696 (2009).
- Anderson, T. J. *et al.* Are transporter genes other than the chloroquine resistance locus (*pfcr*) and multidrug resistance gene (*pfmdr*) associated with antimalarial drug resistance? *Antimicrob. Agents Chemother.* **49**, 2180–2188 (2005).
- Jambou, R. *et al.* Resistance of *Plasmodium falciparum* field isolates to in-vitro artemether and point mutations of the SERCA-type PfATPase6. *Lancet* **366**, 1960–1963 (2005).
- Krishna, S. *et al.* Artemisinins and the biological basis for the PfATP6/SERCA hypothesis. *Trends Parasitol.* **26**, 517–523 (2010).
- Hunt, P. *et al.* Gene encoding a deubiquitinating enzyme is mutated in artesunate- and chloroquine-resistant rodent malaria parasites. *Mol. Microbiol.* **65**, 27–40 (2007).
- Hunt, P. *et al.* Experimental evolution, genetic analysis and genome re-sequencing reveal the mutation conferring artemisinin resistance in an isogenic lineage of malaria parasites. *BMC Genomics* **11**, 499 (2010).
- Borges, S. *et al.* Genome-wide scan reveals amplification of *mdr1* as a common denominator of resistance to mefloquine, lumefantrine, and artemisinin in *Plasmodium chabaudi* malaria parasites. *Antimicrob. Agents Chemother.* **55**, 4858–4865 (2011).
- Chavchich, M. *et al.* Role of *pfmdr1* amplification and expression in induction of resistance to artemisinin derivatives in *Plasmodium falciparum*. *Antimicrob. Agents Chemother.* **54**, 2455–2464 (2010).
- Chen, N. *et al.* Deamplification of *pfmdr1*-containing amplicon on chromosome 5 in *Plasmodium falciparum* is associated with reduced resistance to artemisinin acid in vitro. *Antimicrob. Agents Chemother.* **54**, 3395–3401 (2010).
- Picot, S. *et al.* A systematic review and meta-analysis of evidence for correlation between molecular markers of parasite resistance and treatment outcome in falciparum malaria. *Malar. J.* **8**, 89 (2009).
- Price, R. N. *et al.* Mefloquine resistance in *Plasmodium falciparum* and increased *pfmdr1* gene copy number. *Lancet* **364**, 438–447 (2004).
- Sidhu, A. B. *et al.* Decreasing *pfmdr1* copy number in *Plasmodium falciparum* malaria heightens susceptibility to mefloquine, lumefantrine, halofantrine, quinine, and artemisinin. *J. Infect. Dis.* **194**, 528–535 (2006).
- Yuan, J. *et al.* Chemical genomic profiling for antimalarial therapies, response signatures, and molecular targets. *Science* **333**, 724–729 (2011).
- Amambua-Ngwa, A. *et al.* Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. *PLoS Genet.* **8**, e1002992 (2012).
- Adams, J. *et al.* The kelch repeat superfamily of proteins: propellers of cell function. *Trends Cell Biol.* **10**, 17–24 (2000).
- Prag, S. & Adams, J. C. Molecular phylogeny of the kelch-repeat superfamily reveals an expansion of BTB/kelch proteins in animals. *BMC Bioinformatics* **4**, 42 (2003).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the patients and field staff involved in clinical trials and sample collections. We are grateful to the provincial health department directors and other staff of the Cambodian Ministry of Health. Clinical trials and sample collections were supported in part by the Global Fund Grant Malaria Program Rounds 6 (CAM-607-G10M-CNM3) and 9 (CAM-S10-G14-M), the Bill and Melinda Gates Foundation and USAID (through the World Health Organization), the US DOD Global Epidemic Information System, and the Intramural Research Program, NIAID, NIH. Laboratory work was supported by grants from Banque Natixis (to O.M.-P. and D.M.) and Laboratoire d'Excellence IBEID (Agence Nationale de la Recherche, France) and Institut Pasteur, Division International (ACIP A-10-2010). B.W. was supported by a postdoctoral fellowship from Institut Pasteur, Division International; J.B. by an Institut Pasteur Paris Master-Pro fellowship; and D.M. by the French Ministry of Foreign Affairs. We are grateful to the Wellcome Trust Sanger Institute and the MalariaGen resource centre for sequencing, genotyping and population structure analysis of some Cambodian clinical samples, funded by the Wellcome Trust (098051; 090770/Z/

09/Z) and the MRC (G0600718). We thank the Rotary Club-Versailles for funding computer equipment. P.R., D.M.B. and W.O.R. are staff members of the World Health Organization and the US Navy, respectively. They alone are responsible for the views expressed in this publication, and they do not necessarily represent the decisions, policy or views of the World Health Organization or the US Navy.

**Author Contributions** B.W., S.M., A.B. and F.B.-V. produced the F32-ART5 and F32-TEM clonal lines and analysed their survival rates. F.A. and J.B. developed computational components of the whole-genome sequence analysis. C.B. and L.M. performed whole-genome sequencing. F.A., C.A., S.K., V.D., P.L., R.L., S.D., Se.S., So.S., C.M.C., D.M.B., W.O.R., B.G., T.F., P.R., J.L.B., R.M.F. and D.M. conducted clinical studies and collected parasite isolates. A.-C.L., N.K., S.K., V.D., S.M. and A.B. performed PCR and sequencing analyses. B.W., F.B.-V., V.D. and D.M. performed *in vitro* assays (RSA<sub>0-3</sub>h). O.M. provided genotyping and population structure data for Cambodian parasite isolates. J.-C.B. and O.M.-P. performed three-dimensional structure modelling. F.A., R.M.F., F.B.-V., O.M.-P.

and D.M. conceived of the study, supervised the project, processed the data and wrote the manuscript with contributions from B.W., C.A., A.B. and J.-C.B.

**Author Information** The following reagents have been deposited to the MR4/BEI by D.M.: MRA-1236 (*Plasmodium falciparum* IPC 3445 Pailin Cambodia 2010), MRA-1237 (*Plasmodium falciparum* IPC 3663 Pailin Cambodia 2010), MRA-1238 (*Plasmodium falciparum* IPC 4884 Pursat Cambodia 2011), MRA-1239 (*Plasmodium falciparum* IPC 5188 Ratanakiri Cambodia 2011), MRA-1240 (*Plasmodium falciparum* IPC 5202 Battambang Cambodia 2011) and MRA-1241 (*Plasmodium falciparum* IPC 4912 Monduliri Cambodia 2011). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.A. ([frederic.ariet@pasteur.fr](mailto:frederic.ariet@pasteur.fr)), O.M.-P. ([odile.puijalon@pasteur.fr](mailto:odile.puijalon@pasteur.fr)), D.M. ([dmenard@pasteur-kh.org](mailto:dmenard@pasteur-kh.org)), F.B.-V. ([Francoise.Vical@inserm.fr](mailto:Francoise.Vical@inserm.fr)) and R.M.F. ([rfairhurst@niaid.nih.gov](mailto:rfairhurst@niaid.nih.gov)).



## METHODS

**Artemisinin- and mock-pressured *P. falciparum* F32 lineages.** Mycoplasma-free F32-Tanzania parasites were maintained in human type O red blood cells (RBCs) (Etablissement Français du Sang) diluted to 2.5% haematocrit in RPMI-1640 medium (Invitrogen, San Diego, CA) supplemented with 5% human serum. Parasite cultures were maintained at 37 °C in an atmosphere of 5% CO<sub>2</sub>. Parasitaemia was checked daily and maintained below 10%. For the selection of ART-resistant parasites, asynchronous cultures were adjusted to 5–7% parasitaemia and grown in the presence of escalating doses of artemisinin (from 10 nM to 9 µM) for 24 h for the first 3 years of drug pressure<sup>18</sup>. In the subsequent 2 years, each drug-pressure cycle was done for 48 h with doses ranging from 9 µM to 18 µM. After drug exposure, the medium was discarded and replaced by human-serum-supplemented (20%) drug-free medium. Parasitaemia was monitored daily until it reached 5%. At that time, drug pressure was reapplied. The parasite line obtained after an effective 5 years of discontinuous ART pressure was named F32-ART5. In parallel, the parental F32-Tanzania line was kept as a control in continuous culture for the same time under the same conditions (that is, RBCs, serum and media) but without artemisinin exposure. The resulting control line was called F32-TEM.

**Laboratory-adapted *P. falciparum* lines.** Reference DNA was extracted from the laboratory-adapted *P. falciparum* lines 3D7 (MR4, Manassas, VA), 89F5 Palo Alto Uganda (a clone from the Palo Alto line, originating from Uganda in 1978, which displayed high susceptibility to artemether treatment in the *Saimiri sciureus* experimental host (O. Mercereau-Puijalon, H. Contamin and J.-C. Barale, unpublished data)) and K1992, an isolate collected in Pailin in 1992 before the mass deployment of ART in that area (provided by the French National Reference Center of Malaria). Parasite DNA was extracted from frozen blood aliquots (200 µl) using the Mini blood kit (Qiagen) according to the manufacturer's instructions.

**Culture-adapted *P. falciparum* isolates from Cambodia.** Fifty clinical *P. falciparum* isolates from Cambodia (collected in 2010 and 2011) were adapted to *in vitro* culture as described in ref. 45. Their geographic origin is indicated in Supplementary Table 1. Parasite clearance rates were not determined for these patient isolates, as they were collected during field trials that did not include such measurements. Parasite DNA was extracted from frozen blood aliquots (200 µl) using the Mini blood kit (Qiagen).

**Ring-stage survival assay.** The ring-stage survival assay (RSA<sub>0–3h</sub>) was carried out as described in ref. 13 using highly synchronous parasite cultures. In brief, 0–3 h post-invasion ring-stage parasites were exposed to 700 nM DHA (dihydroartemisinin, obtained from WWARN (<http://www.wwarn.org/research/tools/qaqc>)) or its solvent DMSO for 6 h, washed and then cultivated for the next 66 h without drug. Survival rates were assessed microscopically by counting in Giemsa-stained thin smears the proportion of viable parasites that developed into second-generation rings or trophozoites with normal morphology.

**Ethical clearance.** Ethical clearances for the collection of parasite isolates from patients were obtained from the Cambodian National Ethics Committee for Health Research, the Institutional Review Board of the Naval Medical Research Center, the Technical Review Group of the WHO Regional Office for the Western Pacific, and the Institutional Review Board of the National Institute of Allergy and Infectious Diseases. Work was conducted in compliance with all relevant ethical standards and regulations governing research involving human subjects. Written informed consent was obtained from all adult participants or the parents or guardians of children.

**Temporal and geographical sample collection.** A set of 941 clinical *P. falciparum* isolates was collected from patients participating in therapeutic efficacy studies of ACTs, conducted as part of the routine antimalarial drug efficacy monitoring of Cambodia's National Malaria Control Program from 2001 to 2012, and from studies conducted by NAMRU-2 (Extended Data Table 4). Venous blood samples (5 ml) collected in EDTA or ACD were transported to Institut Pasteur du Cambodge in Phnom Penh within 48 h of collection at 4 °C and then kept at –20 °C until DNA extraction. Parasite DNA was extracted from frozen blood aliquots (200 µl) using the Mini blood kit (Qiagen).

**Measurement of parasite clearance half-life.** Patients with uncomplicated or severe *P. falciparum* malaria and initial parasite density  $\geq 10,000 \mu\text{l}^{-1}$  were enrolled in Pursat and Ratanakiri provinces in 2009 and 2010 as described<sup>6,13</sup>. Patients were treated with an ART and their parasite density measured every 6 h from thick blood films until parasitaemia was undetectable. The parasite clearance half-life in 163 patients was derived from these parasite counts using WWARN's on-line Parasite Clearance Estimator (<http://www.wwarn.org/toolkit/data-management/parasite-clearance-estimator>). The study is registered at ClinicalTrials.gov (number NCT00341003).

**Whole-genome sequencing of parasite DNA.** Whole-genome sequencing was performed on F32-Tanzania, F32-TEM, the F32-ART5 lineage (4 time-points), three reference strains (3D7, 89F5 and K1992) and 21 parasite isolates from Cambodia, using an Illumina paired-reads sequencing technology. Illumina library preparation

and sequencing followed standard protocols developed by the supplier. Briefly, genomic DNA was sheared by nebulization, and sheared fragments were end-repaired and phosphorylated. Blunt-end fragments were A-tailed, and sequencing adapters were ligated to the fragments. Inserts were sized using Agencourt AMPure XP Beads ( $\pm 500$  bp; Beckman Coulter Genomics) and enriched using 10 cycles of PCR before library quantification and validation. Hybridization of the library to the flow cell and bridge amplification was performed to generate clusters, and paired-end reads of 100 cycles were collected on a HiSeq 2000 instrument (Illumina). After sequencing was complete, image analysis, base calling and error estimation were performed using Illumina Analysis Pipeline version 1.7.

Raw sequence files were filtered using Fqquality tool, a read-quality filtering software developed by N. Joly, which enables the trimming of the first and last low-quality bases in reads. The trimmed reads from controlled Fastq files were mapped on a reference genome (*P. falciparum* 3D7) with the Burrows-Wheeler Alignment (BWA), generating a BAM file (a binary file of tab-delimited format SAM). Next, we used Samtools to prepare a pileup file, which was formatted using in-house software to implement the data into the Wholegenome Data Manager (WDM) database (Beghain *et al.*, in preparation). WDM software is designed to compare and/or align partial or whole *P. falciparum* genomes.

**Sequencing genes containing non-synonymous SNPs in F32-ART5.** PCR amplification of selected genes was performed using the primers listed in Extended Data Table 1. Two µl of DNA was amplified with 1 µM of each primer, 0.2 mM dNTP (Solis Biodyne), 3 mM MgCl<sub>2</sub> and 2 U *Taq* DNA polymerase (Solis Biodyne), using the following cycling program: 5 min at 94 °C, then 40 cycles of 30 s at 94 °C, 90 s at 60 °C, 90 s at 72 °C and final extension 10 min at 72 °C. PCR products were detected by 2% agarose gel electrophoresis and ethidium bromide staining. Double-strand sequencing of PCR products was performed by Beckman Coulter Genomics. Sequences were analysed with MEGA 5 software version 5.10 in order to identify specific SNP combinations.

**Sequencing the K13-propeller domain.** The K13-propeller domain was amplified using the following primers: for the primary PCR (K13-1 5'-cggagtgaccaatctggga-3' and K13-4 5'-gggaatctgtgtggaacacg-3') and the nested PCR (K13-2 5'-gccaaagtcg cattattg-3' and K13-3 5'-gccttggtgaagaagcaga-3'). One µl of DNA was amplified with 1 µM of each primer, 0.2 mM dNTP (Solis Biodyne), 3 mM MgCl<sub>2</sub> and 2 U *Taq* DNA polymerase (Solis Biodyne), using the following cycling program: 5 min at 94 °C, then 40 cycles of 30 s at 94 °C, 90 s at 60 °C, 90 s at 72 °C and final extension 10 min at 72 °C. For the nested PCR, 2 µl of primary PCR products were amplified under the same conditions, except for the MgCl<sub>2</sub> concentration (2.5 mM). PCR products were detected using 2% agarose gel electrophoresis and ethidium bromide staining. Double-strand sequencing of PCR products was performed by Macrogen. Sequences were analysed with MEGA 5 software version 5.10 to identify specific SNP combinations.

**Deep-sequencing of clinical parasite isolates and population structure analysis.** DNA extraction, Illumina sequencing and SNP genotyping of clinical parasite isolates obtained from malaria patients in Pursat and Ratanakiri provinces, Cambodia, have been previously described<sup>15</sup>. Population structure analysis of these parasites identified four subpopulations: KH1, KH2, KH3 and KH4. Parasites with <80% ancestry from any of these four groups were deemed admixed (KHA).

**Temporal acquisition of mutations in the F32-ART5 lineage.** The F32-ART5 lineage was explored by whole-genome sequencing using samples collected at time 0 (original F32-Tanzania clonal line), day 196 (0.2-µM pressure cycle no. 23), day 385 (1.8-µM pressure cycle no. 39), day 618 (9-µM pressure cycle no. 56) and day 2,250 (9-µM pressure cycle no. 120). The F32-TEM sample was collected on day 2,250. Additional samples collected at the time of the 30th, 33rd, 34th, 36th, 68th and 98th pressure cycles were studied by PCR. DNA from parasite cultures was extracted using the High Pure PCR Template Preparation Kit (Roche Diagnostics) according to the manufacturer's instructions.

The F32-ART5 samples tested in the ring-stage survival assay (RSA<sub>0–3h</sub>) were collected at the time of the 17th, 48th and 122nd pressure cycles (0.04, 2.7 and 9 µM ART), respectively. The F32-TEM sample was collected at the last mock pressure cycle. The RSA<sub>0–3h</sub> survival rates were determined in triplicate experiments with different batches of red blood cells, and evaluated as above using Giemsa-stained thin smears read by two independent microscopists (B.W. and F.B.-V.). Survival rates were compared using Mann-Whitney *U* test. The RSA<sub>0–3h</sub> survival rates of the F32-ART5 samples were as follows: at drug-pressure cycles: no. 17 ( $n = 3$ , median 0%, IQR 0–0.07), no. 48 ( $n = 3$ , median 11.7%, IQR 10.3–14.6;  $P = 0.04$  for no. 17 versus no. 48, Mann-Whitney *U* test) and no. 122 ( $n = 3$ , median 12.8%, IQR 10.6–14.5,  $P = 0.04$  and  $P = 0.82$  for no. 17 versus no. 122 and no. 48 versus no. 122). The RSA<sub>0–3h</sub> survival rate of the F32-TEM line was also determined in triplicate experiments ( $n = 3$ , median 0%, IQR 0–0.05,  $P = 0.81$  for TEM versus no. 17, Mann-Whitney *U* test).

**Prevalence of K13-propeller mutations in 886 clinical parasite isolates collected in six Cambodian provinces in 2001–2012.** The K13-propeller was genotyped by

sequencing PCR products amplified from 886 archived blood samples. The number of samples analysed from each province each year is indicated in Fig. 3. Fisher's exact test was used to compare the frequency of isolates harbouring a wild-type K13-propeller sequence in each province over time. A significant decrease of the frequency of the wild-type K13-propeller allele was observed in the western provinces during the decade. In Pailin, it decreased from 30.0% in 2001–2002 (12/40) to 4.8% in 2011–2012 (4/84),  $P = 0.0002$ , in Battambang from 71.9% in 2001–2002 (46/64) to 7.0% in 2011–2012 (5/71),  $P < 10^{-6}$ , in Pursat from 50.0% in 2003–2004 (5/10) to 10.5% in 2011–2012 (2/19),  $P = 0.03$ ; and in Kratie from 93.3% in 2001–2002 (14/15) to 29.4% in 2011–2012 (5/17),  $P = 0.0003$ . Significant decreases in wild-type allele frequency were not observed in Preah Vihear (from 92.6% in 2001–2002 (25/27) to 84.2% in 2011–2012 (16/19),  $P = 0.63$ ); or Ratanakiri (from 96.4% in 2001–2002 (54/56) to 94.3% in 2011–2012 (33/35),  $P = 0.63$ ). The frequency of C580Y increased in Pailin from 45.0% (18/40) in 2001–2002 to 88.1% (74/84) in 2011–2012 ( $P < 10^{-6}$ ), and in Battambang from 7.8% (5/64) in 2001–2002 to 87.3% (62/71) in 2011–2012 ( $P < 10^{-6}$ ) indicating its rapid invasion of the population and near fixation in these provinces.

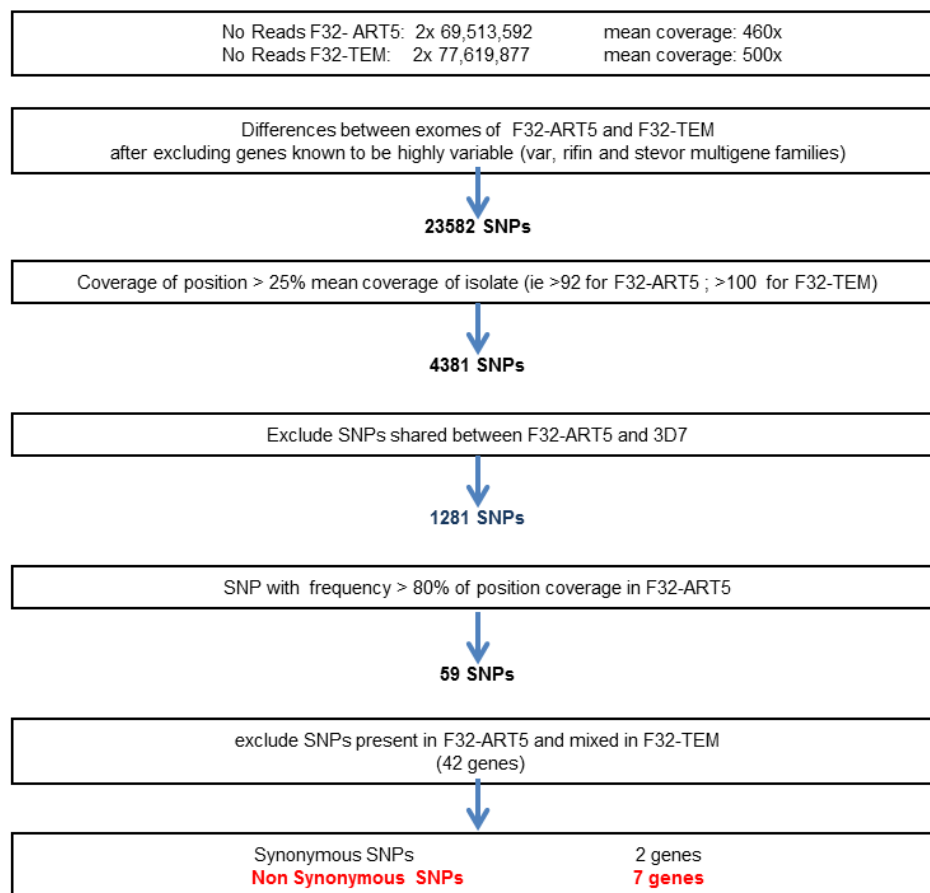
**Three-dimensional structure modelling of the K13-propeller.** The 3D-structural model of the kelch propeller domain of PF3D7\_1343700 ('K13-propeller') was obtained by homology modelling satisfying spatial restraints using Modeller v9.11 (<http://modbase.compbio.ucsf.edu>). The 295 amino acids composing the K13-propeller are 22%, 25% and 25% identical to the kelch propeller domain of the human KEAP1 (Protein Data Bank (PDB; <http://www.rcsb.org/>) 2FLU), KLHL12 (PDB 2VPJ) and KLHL2 (PDB 2XN4) proteins, respectively, that were used as templates to model the 3D-structure of the K13-propeller. The reliability of the obtained model was assessed using two classical criteria. First, the significance of the sequence alignment between the K13-propeller domain and one template was confirmed by an  $E$ -value = 0, as calculated by Modeller using the Built-Profile routine. Second, the model achieved a GA341 model score = 1 (a score  $\geq 0.7$  corresponds to highly reliable models). Localization of the mutants in the K13-propeller 3D-model was prepared using the PyMOL Molecular Graphics System, version 1.5.0.4 (Schrödinger; <http://www.pymol.org>).

**Statistical analysis.** Data were analysed with Microsoft Excel and MedCalc version 12. Quantitative data were expressed as median, interquartile range (IQR). The Mann–Whitney  $U$  test (independent samples, two-sided) was used to compare two groups, and the Kruskal–Wallis test ( $H$ -test, two-sided) was used to compare more than two groups. The Spearman's rho rank correlation coefficient (and the 95% confidence interval for the correlation coefficient) was used to measure the strength of relationship between the prevalence of wild-type K13-propeller allele and the frequency of day 3 positivity (defined as persistence of microscopically detectable parasites on the third day of artemisinin-based combination therapy)<sup>2</sup>. Fisher's exact test was used to compare frequency data and the Clopper–Pearson exact method based on the beta distribution was used to determine

95% confidence intervals for proportions. Differences were considered statistically significant when  $P$  values were less than 0.05.

45. Witkowski, B. *et al.* Reduced artemisinin susceptibility of *Plasmodium falciparum* ring stages in western Cambodia. *Antimicrob. Agents Chemother.* **57**, 914–923 (2013).
46. Padmanabhan, B. *et al.* Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. *Mol. Cell* **21**, 689–700 (2006).
47. Boyden, L. M. *et al.* Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities. *Nature* **482**, 98–102 (2012).
48. Li, X., Zhang, D., Hannink, M. & Beamer, L. J. Crystal structure of the Kelch domain of human Keap1. *J. Biol. Chem.* **279**, 54750–54758 (2004).
49. Itoh, K. *et al.* Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. *Genes Dev.* **13**, 76–86 (1999).
50. Zhang, D. D. & Hannink, M. Distinct cysteine residues in Keap1 are required for Keap1-dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and oxidative stress. *Mol. Cell. Biol.* **23**, 8137–8152 (2003).
51. Bozdech, Z. & Ginsburg, H. Antioxidant defense in *Plasmodium falciparum*—data mining of the transcriptome. *Malar. J.* **3**, 23 (2004).
52. Nesser, N. K., Peterson, D. O. & Hawley, D. K. RNA polymerase II subunit Rpb9 is important for transcriptional fidelity *in vivo*. *Proc. Natl Acad. Sci. USA* **103**, 3268–3273 (2006).
53. Kettenberger, H., Armache, K. J. & Cramer, P. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114**, 347–357 (2003).
54. Dorin-Semblat, D., Sicard, A., Doerig, C., Ranford-Cartwright, L. & Doerig, C. Disruption of the PfPK7 gene impairs schizogony and sporogony in the human malaria parasite *Plasmodium falciparum*. *Eukaryot. Cell* **7**, 279–285 (2008).
55. Tewari, R. *et al.* The systematic functional analysis of *Plasmodium* protein kinases identifies essential regulators of mosquito transmission. *Cell Host Microbe* **8**, 377–387 (2010).
56. Rosenthal, P. J., McKerrow, J. H., Aikawa, M., Nagasawa, H. & Leech, J. H. A malarial cysteine proteinase is necessary for hemoglobin degradation by *Plasmodium falciparum*. *J. Clin. Invest.* **82**, 1560–1566 (1988).
57. Sijwali, P. S. *et al.* *Plasmodium falciparum* cysteine protease falcipain-1 is not essential in erythrocytic stage malaria parasites. *Proc. Natl Acad. Sci. USA* **101**, 8721–8726 (2004).
58. Sijwali, P. S., Koo, J., Singh, N. & Rosenthal, P. J. Gene disruptions demonstrate independent roles for the four falcipain cysteine proteases of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **150**, 96–106 (2006).
59. Klonis, N. *et al.* Altered temporal response of malaria parasites determines differential sensitivity to artemisinin. *Proc. Natl Acad. Sci. USA* **110**, 5157–5162 (2013).
60. Lobo, C. A., Fujioka, H., Aikawa, M. & Kumar, N. Disruption of the *Pf*g27 locus by homologous recombination leads to loss of the sexual phenotype in *P. falciparum*. *Mol. Cell* **3**, 793–798 (1999).
61. Olivieri, A. *et al.* The *Plasmodium falciparum* protein Pf27 is dispensable for gametocyte and gamete production, but contributes to cell integrity during gametocytogenesis. *Mol. Microbiol.* **73**, 180–193 (2009).
62. Sharma, A., Sharma, I., Kogkasuriyachai, D. & Kumar, N. Structure of a gametocyte protein essential for sexual development in *Plasmodium falciparum*. *Nature Struct. Biol.* **10**, 197–203 (2003).

**a**

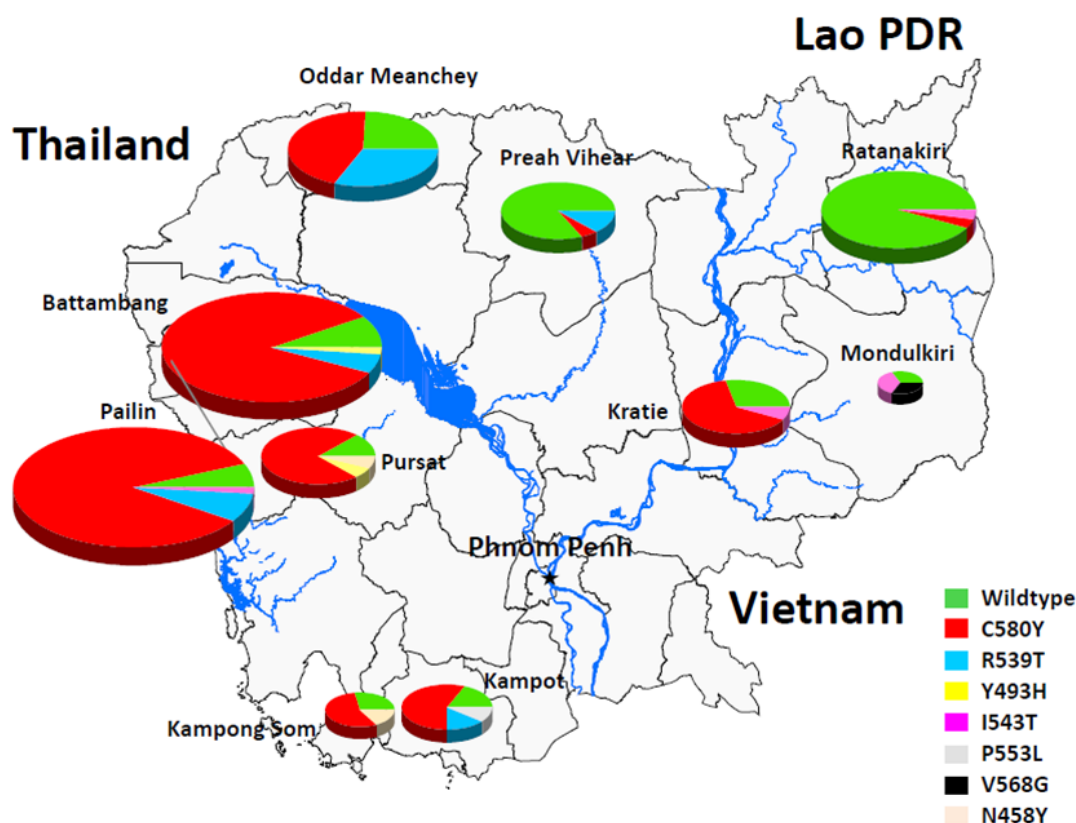


**b**

Chromosome	Position	Gene ID	F32-ART5			F32-TEM		
			coverage	No reads with mutant SNP	% reads with mutant SNP	coverage	No reads with w.t. SNP	% reads with w.t. SNP
1	394452	PF3D7_0110400;DNA-directed RNA polymerase 2, putative	224	222	99.11	335	334	99.70
2	542625	PF3D7_0213400;protein kinase 7 (PK7)	242	242	100.00	403	403	100.00
11	593379	PF3D7_1115700;cysteine proteinase falcipain 2a	234	231	98.72	290	289	99.66
13	121689	PF3D7_1302100;gamete antigen 27/25 (Pfg27)	261	259	99.23	343	342	99.71
13	1725570	PF3D7_1343700;kelch protein, putative	1004	1004	100.00	1161	1160	99.91
14	2442240	PF3D7_1459600;conserved Plasmodium protein, unknown	165	142	86.06	225	225	100.00
14	2612177	PF3D7_1464500;conserved Plasmodium membrane protein, unknown	401	399	99.50	428	428	100.00

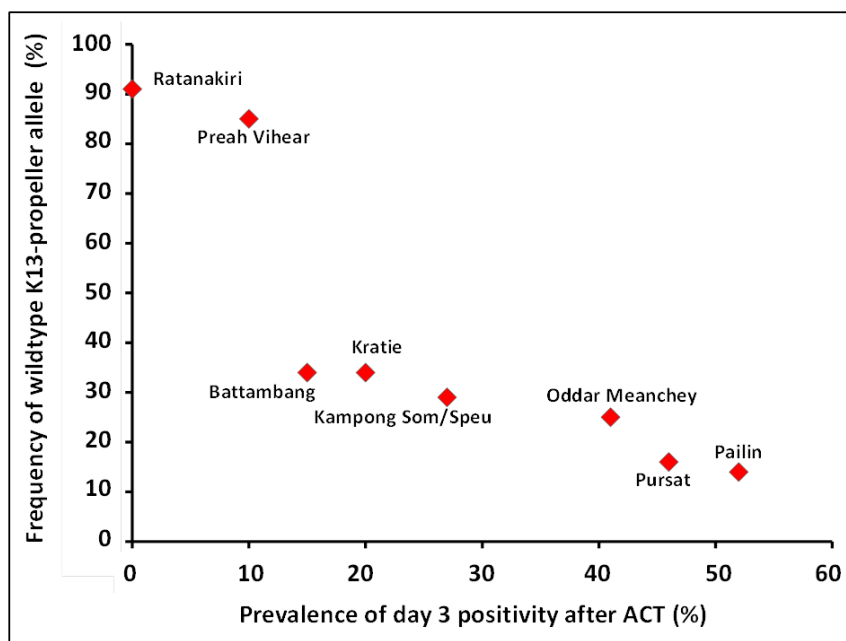
**Extended Data Figure 1 | SNP-calling algorithm and sequence and coverage of SNPs. a**, SNP-calling algorithm of the whole-genome sequence comparison of F32-ART5 and F32-TEM. **b**, Sequence and coverage of SNPs in seven candidate genes differing in F32-TEM and F32 ART5.





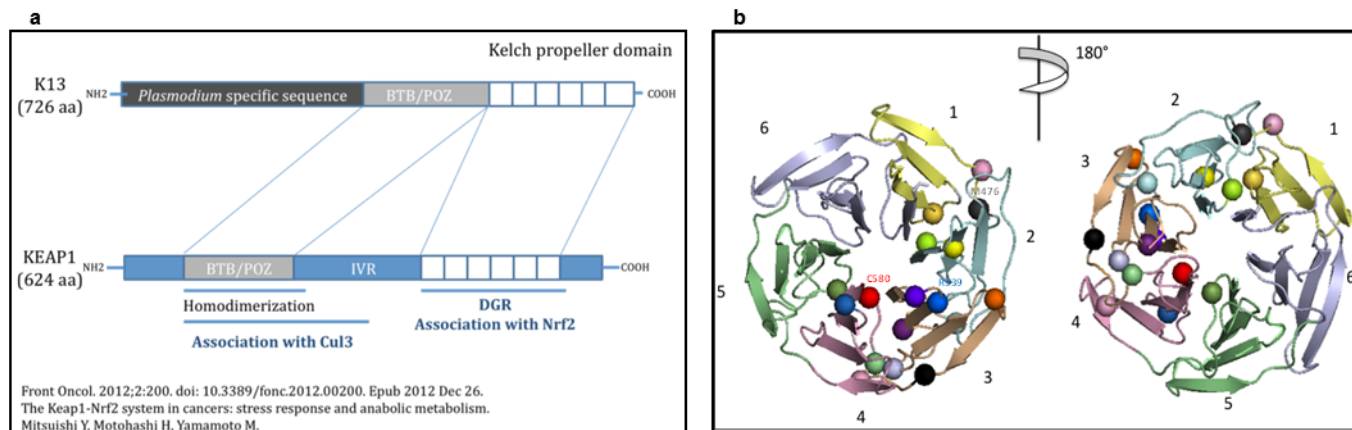
**Extended Data Figure 2 | Geographic distribution of K13-propeller alleles in Cambodia in 2011–2012.** Pie charts show K13-propeller allele frequencies among 300 parasite isolates in ten Cambodian provinces. Pie sizes are proportional to the number of isolates and the different alleles are colour-coded as indicated. The frequencies (95% confidence interval) of mutant K13-

propeller alleles are: Pailin (95%, 88–99,  $n = 84$ ), Battambang (93%, 87–99,  $n = 71$ ), Pursat (89%, 67–99,  $n = 19$ ), Kampong Som (83%, 52–98,  $n = 12$ ), Kampong Som (71%, 29–96,  $n = 7$ ), Oddar Meanchey (76%, 58–89,  $n = 33$ ), Preah Vihear (16%, 3–40,  $n = 19$ ), Kratie (71%, 44–90,  $n = 17$ ), Mondulakiri (67%, 9–99,  $n = 3$ ) and Ratanakiri (6%, 1–19,  $n = 35$ ).



**Extended Data Figure 3 | Correlation between the frequency of wild-type K13-propeller alleles and the prevalence of day 3 positivity after ACT treatment in eight Cambodian provinces.** The frequency of day 3 positivity is plotted against the frequency of wild-type K13-propeller alleles. Data are derived from patients treated with an ACT for *P. falciparum* malaria in 2010–2012 in eight Cambodian provinces (Extended Data Figure 2): Pailin ( $n = 86$ , 2011 WHO therapeutic efficacy study, artesunate-mefloquine); Pursat ( $n = 32$ , 2012 WHO therapeutic efficacy study, dihydroartemisinin-piperaquine); Oddar Meanchey ( $n = 32$ , 2010 NAMRU-2 therapeutic efficacy study,

artesunate-mefloquine); Kampong Som/Speu ( $n = 7$ , 2012 WHO therapeutic efficacy study, dihydroartemisinin-piperaquine); Battambang ( $n = 18$ , 2012 WHO therapeutic efficacy study, dihydroartemisinin-piperaquine); Kratie ( $n = 15$ , 2011 WHO therapeutic efficacy study, dihydroartemisinin-piperaquine); Preah Vihear ( $n = 19$ , 2011 WHO therapeutic efficacy study, dihydroartemisinin-piperaquine); Ratanakiri ( $n = 32$ , 2010 WHO therapeutic efficacy study, dihydroartemisinin-piperaquine). Spearman's coefficient of rank correlation (8 sites):  $r = -0.99$ , 95% confidence interval  $-0.99$  to  $-0.96$ ,  $P < 0.0001$ .



**Extended Data Figure 4 | Schematic representation of homology between *P. falciparum* K13 and human KEAP1 proteins and structural 3D model of the K13-propeller domain.** **a**, Schematic representation of the predicted *PF3D7\_1343700* protein and homology to human KEAP1. Similar to KEAP1, *PF3D7\_1343700* contains a BTB/POZ domain and a C-terminal 6-blade propeller, which assembles kelch motifs consisting of four anti-parallel beta sheets. **b**, Structural 3D model of the K13-propeller domain showing the six kelch blades numbered 1 to 6 from N to C terminus and colour-coded as in Supplementary Fig. 1. The level of amino-acid identity between the K13-propeller and kelch domains of proteins with solved 3D structures, including human KEAP1<sup>46,47</sup>, enabled us to model the 3D structure of the K13-propeller and to map the mutations selected under ART pressure (Extended data Table 5). The accuracy of the K13-propeller 3D model was confirmed by Modeller-specific model/fold criteria of reliability (see Methods). We predict that the K13-propeller folds into a 6-bladed  $\beta$ -propeller structure<sup>48</sup> closed by the interaction between a C-terminal beta-sheet and the N-terminal blade<sup>46,48</sup>. The first domain has three  $\beta$ -sheets, the fourth one being contributed by an extra C-terminal  $\beta$ -sheet called  $\beta'$ 1 in Supplementary Fig. 1. The human KEAP1 kelch propeller scaffold is destabilized by a variety of mutations affecting intra- or inter-blade interactions in human lung cancer<sup>46</sup> and hypertension<sup>47</sup>. The positions of the various mutations are indicated by a

sphere, colour-coded as in Figs 2–4. The M476 residue mutated in F32-ART5 is indicated in dark grey. Like the mutations observed in human KEAP1<sup>46,47</sup>, many K13-propeller mutations are predicted to alter the structure of the propeller or modify surface charges, and as a consequence alter the biological function of the protein. Importantly, the two major mutations C580Y (red) and R539T (blue) observed in Cambodia are both non-conservative and located in organized secondary structures: a  $\beta$ -sheet of blade 4 where it is predicted to alter the integrity of this scaffold and at the surface of blade 3, respectively. The kelch propeller domain of KEAP1 is involved in protein–protein interactions like most kelch containing modules<sup>43</sup>. KEAP1 is a negative regulator of the inducible Nrf2-dependent cytoprotective response, sequestering Nrf2 in the cytoplasm under steady state. Upon oxidative stress, the Nrf2/KEAP1 complex is disrupted, and Nrf2 translocates to the nucleus, where it induces transcription of cytoprotective ARE-dependent genes<sup>49,50</sup>. We speculate that similar functions may be performed by *PF3D7\_1343700* in *P. falciparum*, such that mutations of the K13-propeller impair its interactions with an unknown protein partner, resulting in a deregulated anti-oxidant/cytoprotective response. The *P. falciparum* anti-oxidant response is maximal during the late trophozoite stage, when haemoglobin digestion and metabolism are highest<sup>51</sup>. Its regulation is still poorly understood and no Nrf2 orthologue could be identified in the *Plasmodium* genome.



**Extended Data Table 1 | Sequence of the primers used to amplify the genes containing nonsynonymous single-nucleotide polymorphisms in F32-ART5**

Targeted gene	Primer forward sequence	Primer reverse sequence
PF3D7_0110400	5'-ttgagcttctttttcccaataatggc-3'	5'-tgatatagtttgtaggagctgtgag-3'
PF3D7_0213400	5'-gtgaaaaggataataaattctatgcc-3'	5'-tatctaccatataattctgattctcc-3'
PF3D7_1115700	5'-agcaagaacgttttgtgtaaa-3'	5'-gaattctttaatggttttgaagat-3'
PF3D7_1302100	5'-taatatgtaaagtgattatgtatatogo-3'	5'-atgctagagaagttaagagaagaagcg-3'
PF3D7_1343700	5'-agaagagccatcatatccccc-3'	5'-agtggaagacatcatgtaaccag-3'
PF3D7_1459600	5'-atatgagtaaaatgtcaggttttgg-3'	5'-tgcttggtgtgattcatgggg-3'
PF3D7_1464500	5'-aaatagttggcgtagctcag-3'	5'-tatacaattaagtgtatcacaacg-3'

**Extended Data Table 2 | Description of the eight nonsynonymous, single-nucleotide polymorphisms acquired in the F32-ART5 compared to the F32-TEM lineage during an effective 5-year discontinuous exposure to increasing concentrations of artemisinin**

Gene ID (Plasmodb 9.1)	Annotation	Chromosome/ position mutated	Nucleotide position in coding sequence	F32- TEM codon <sup>a</sup>	Codon F32-ART5 lineage					
					Day 0	Drug pressure cycle #				Mutant codon
						23	39	56	120	
					-	0.2 $\mu$ M ART <sup>*</sup>	1.8 $\mu$ M ART <sup>*</sup>	9 $\mu$ M ART <sup>*</sup>	9 $\mu$ M ART <sup>*</sup>	
PF3D7_0110400	DNA-directed RNA polymerase 2 complex subunit RPB9, putative	01/39452	173	gAt	gAt	gAt	<b>gTg</b>	<b>gTg</b>	<b>gTg</b>	D56V
PF3D7_1343700 <sup>a</sup>	kelch protein, putative, called here 'K13'	13/1725570	1428	atG	atG	atG	<b>atA</b>	<b>atA</b>	<b>atA</b>	M476I
PF3D7_0213400	protein kinase 7 (PK7)	02/542625 02/542627	310 312	GaA	GaA	GaA	GaA	GaA	<b>TaG</b>	E104stop
PF3D7_1115700	cysteine proteinase falcipain 2a	11/593378	206	tCa	tCa	tCa	tCa	tCa	<b>tGa</b>	S69stop
PF3D7_1302100	gamete antigen 27/25 (Pfg27)	13/121689	601	Cca	Cca	Cca	Cca	Cca	<b>Aca</b>	P201T
PF3D7_1459600 <sup>a</sup>	conserved <i>Plasmodium</i> protein, unknown function	14/2442240	896	aGt	aGt	aGt	aGt	aGt	<b>aCt</b>	S299T
PF3D7_1464500	conserved <i>Plasmodium</i> membrane protein, unknown function	14/2612177	4886	aAt	aAt	aAt	aAt	aAt	<b>aGt</b>	N1629S

<sup>#</sup> 3D7-type sequence; the same codon sequence is also observed in the parental F32-Tanzania line.

<sup>\*</sup> Artemisinin (ART) dose used for selection during the corresponding drug-pressure cycle.

<sup>a</sup> Genes found in the chromosomal location of top-ranked signatures of selection in ref. 16.

## Extended Data Table 3 | Reported characteristics of the genes mutated in F32-ART5 parasites

**PF3D7\_0110400 (PFA0505c)**, is a two-exon gene, codes for the RNA Polymerase II subunit 9 (RPB9), a small integral Pol II subunit, which is highly conserved among eukaryotes. The yeast RPB9 ortholog has been shown to have a role in assuring the fidelity of transcription *in vivo*. Deletion of the gene results in error-prone transcription<sup>52</sup>. The protein has a predicted zinc ribbon domain similar to the zinc ribbon domain of TFIIS (RNA Polymerase II elongation factor) that contains the essential catalytic Asp-Glu dipeptide<sup>53</sup>. Very little is known on the protein in *Plasmodia*, although the gene is expressed and the protein is present in blood stages (www. plasmodb.org). It is difficult to make any prediction on the possible phenotypic consequences of the D56V mutation, which is located in a *Plasmodium*-specific, well-conserved domain.

**PF3D7\_1343700 (PF13\_0238)**, is a one-exon gene (called here K13) that codes for a putative kelch protein. K13 has a predicted 3-domain structure, with an approx. 225 residue long, *Plasmodium*-specific and well conserved N-terminal domain, followed by a BTB/ POZ domain and a 6-blade C-terminal propeller domain formed of canonical kelch motifs<sup>43,48</sup>. Little is known about the protein in malaria parasites. Proteomics data indicate that it is produced by asexual (trophozoites, schizonts, merozoites and rings) and sexual blood stages (gametocytes) of *P. falciparum*, and that it possesses phosphorylated residues in the N-terminal *Plasmodium*-specific domain (www. plasmodb.org). The M476I mutation is located between the first and second blade of the propeller domain.

**PF3D7\_0213400 (PFB0605w)**, is a four-exon gene that codes for protein kinase 7 (PK7) expressed during the asexual blood stage development, in gametocytes and ookinetes. The E104 stop mutation (two SNPs affecting the same codon) observed in F32-ART5 interrupts the gene resulting in a truncated putative translation product lacking more than 2/3 of its sequence. Studies with genetically inactivated parasites have shown that PK7-KO *P. falciparum* parasites have an asexual growth defect due to a reduced number of merozoites per schizont<sup>54</sup>. Furthermore, PK7 is important for mosquito transmission, with a collapsed number of ookinetes in *P. falciparum*<sup>54</sup> and in *P. berghei*, where no sporoblasts and consequently no sporozoites are formed<sup>55</sup>. This transmission defective phenotype is unlikely to survive in the field.

**PF3D7\_1115700 (PF11\_0165)**, is a one-exon gene that codes for falcipain 2a, a cysteine proteinase produced by maturing blood stages (trophozoites and schizonts) and involved in hemoglobin degradation<sup>56</sup>. The S69stop mutation located in the pro-enzyme region precludes expression of an active enzyme by F32-ART5 parasites. Gene inactivation has shown to induce a transient reduction of hemoglobin degradation compensated by expression of other members of the cysteine proteinases family, with minimal impact on growth rate<sup>57,58</sup>. However, *falcipain 2a* is the only gene from the list of seven affected loci that has been associated with the *in vitro* response to artemisinin. Indeed, it has been convincingly shown that inhibition of falcipain2a-dependent hemoglobin digestion by specific inhibitors or by gene inactivation reduced parasite susceptibility to artemisinins<sup>19</sup>. Moreover, ring stages that do not massively digest hemoglobin display a reduced susceptibility to artemisinins<sup>59</sup>.

**PF3D7\_1302100 (PF13\_0011)**, is a one-exon gene that codes for the gamete antigen 27/25 (Pfg27) produced at the onset of gametocytogenesis. The gene is specific to *P. falciparum* and its close relatives such as *P. reichenowi*. This is an abundant, dimeric phosphorylated cytoplasmic protein that binds RNA. The various KO lines generated display conflicting phenotypes some being deficient in gametocytogenesis<sup>60</sup>, while other Pfg27-defective lines undergo unimpaired gametocytogenesis up to stage V, mature gametocytes although absence of Pfg27 is associated with abnormalities in intracellular architecture of gametocytes<sup>61</sup>. The crystal structure shows that the protein forms a dimer, displays a particular RNA binding fold and possesses two Pro-X-X-Pro motifs (known ligands for various domains, including SH3 modules), which combine to form a receptacle for SH3 modules<sup>62</sup>. The P201T mutation is located in the C-terminal ProX-X-Pro motif and predicted to alter the spatial structure of the interaction domain and thus have functional consequences.

**PF3D7\_1459600 (PF14\_0569)**, is a two-exon gene that codes for a 806 residue-long, conserved protein of unknown function. The *P. yoelii* ortholog has been annotated as the CAAT-box DNA binding subunit B. Close orthologs can be found only among the *Plasmodium* species. Proteomics data indicate that the protein is present in asexual (trophozoites, schizonts, merozoites and rings) and sexual (gametocytes) blood stages of *P. falciparum*. A predicted approx. 130 aa-long Interpro domain suggests presence of an N-terminal multi-helical, alpha-alpha 2-layered structural VHS fold, possibly involved in intracellular membrane trafficking. The rest of the coding sequence carries no specific domain signature. The S299T mutation is located within this "unknown" region.

**PF3D7\_1464500 (PF14\_0603)**, is a five-exon gene that codes for a 3251 residue-long protein of unknown function, with 4 predicted trans-membrane domains, but otherwise no specific domain signature. Apart from proteomics data indicating its expression and phosphorylation in schizonts, with possible expression in gametocytes and sporozoites as well, little is known about its putative function. The N1629S mutation is located in the middle of the protein, with unpredictable phenotypic impact.



**Extended Data Table 4 | Geographic origin and year of collection of archived blood samples studied for K13-propeller polymorphism**

Region	Province	Year of collection						Total
		2001-2002	2003-2004	2005-2006	2007-2008	2009-2010	2011-2012	
Western Cambodia	Battambang	64	0	0	0	0	71	135
	Pailin	40	43	46	95	66	84	374
	Pursat	0	10	0	0	43	19	72
Southern Cambodia	Kampot	0	0	0	0	0	12	12
	Kampong Som	0	0	0	0	0	7	7
Northern Cambodia	Oddar Meanchey	0	0	0	0	0	33	33
	Preah Vihear	27	27	25	24	0	19	122
Eastern Cambodia	Kratie	15	0	0	0	0	17	32
	Monduliri	0	0	0	0	0	3	3
	Ratanakiri	56	30	22	0	8	35	151
<b>Total</b>		<b>202</b>	<b>110</b>	<b>93</b>	<b>119</b>	<b>117</b>	<b>300</b>	<b>941</b>

**Extended Data Table 5 |** Polymorphisms observed in the K13-propeller in Cambodian *P. falciparum* isolates collected in 2001–2012 and in The Gambia (ref. 42)

<b>Codon Position</b>	<b>Amino Acid reference</b>	<b>Nucleotide reference</b>	<b>Amino Acid mutation</b>	<b>Nucleotide mutation</b>
449	G	ggt	A	gCt
458	N	aat	Y	Tat
474	T	aca	I	aTa
476*	M	atg	I	atA
481	A	gct	V	gTt
493	Y	tac	H	Cac
508	T	act	N	aAt
527	P	cct	T	Act
533	G	ggt	S	Agt
537	N	aat	I	aTt
539	R	aga	T	aCa
543	I	att	T	aCt
553	P	ccg	L	cTg
561	R	cgt	H	cAt
568	V	gtg	G	gGg
574	P	cct	L	cTt
580	C	tgt	Y	tAt
584	D	gat	V	gTt
612**	E	gaa	D	gaT
623	S	agt	C	Tgt

\* Observed in F32-ART5, not observed in Cambodia

\*\* Reported in The Gambia<sup>42</sup>, not observed in Cambodia

**Extended Data Table 6 | Association between polymorphisms observed in the K13-propeller and KH subpopulations (ref. 15) in 150 *P. falciparum* isolates collected in 2009–2010 in Pursat ( $n = 103$ ) and Ratanakiri ( $n = 47$ ) provinces, Cambodia**

KH group	Province	Mutations in the K13-propeller				Total
		Wildtype	C580Y	R539T	Y493H	
KH1	Pursat	7	0	0	2	9
	Ratanakiri	46	0	0	0	46
KH2	Pursat	0	25	0	1	26
	Ratanakiri	0	0	0	0	0
KH3	Pursat	3	7	4	0	14
	Ratanakiri	0	0	0	0	0
KH4	Pursat	0	0	0	12	12
	Ratanakiri	0	0	0	0	0
KHA	Pursat	15	19	2	6	42
	Ratanakiri	1	0	0	0	1
Total		72	51	6	21	150



# Structural basis for $\text{Ca}^{2+}$ selectivity of a voltage-gated calcium channel

Lin Tang<sup>1,2\*</sup>, Tamer M. Gamal El-Din<sup>1\*</sup>, Jian Payandeh<sup>1†</sup>, Gilbert Q. Martinez<sup>1</sup>, Teresa M. Heard<sup>1</sup>, Todd Scheuer<sup>1</sup>, Ning Zheng<sup>1,2</sup> & William A. Catterall<sup>1</sup>

Voltage-gated calcium ( $\text{Ca}_V$ ) channels catalyse rapid, highly selective influx of  $\text{Ca}^{2+}$  into cells despite a 70-fold higher extracellular concentration of  $\text{Na}^+$ . How  $\text{Ca}_V$  channels solve this fundamental biophysical problem remains unclear. Here we report physiological and crystallographic analyses of a calcium selectivity filter constructed in the homotetrameric bacterial  $\text{Na}_V$  channel  $\text{Na}_V\text{Ab}$ . Our results reveal interactions of hydrated  $\text{Ca}^{2+}$  with two high-affinity  $\text{Ca}^{2+}$ -binding sites followed by a third lower-affinity site that would coordinate  $\text{Ca}^{2+}$  as it moves inward. At the selectivity filter entry, Site 1 is formed by four carboxyl side chains, which have a critical role in determining  $\text{Ca}^{2+}$  selectivity. Four carboxyls plus four backbone carbonyls form Site 2, which is targeted by the blocking cations  $\text{Cd}^{2+}$  and  $\text{Mn}^{2+}$ , with single occupancy. The lower-affinity Site 3 is formed by four backbone carbonyls alone, which mediate exit into the central cavity. This pore architecture suggests a conduction pathway involving transitions between two main states with one or two hydrated  $\text{Ca}^{2+}$  ions bound in the selectivity filter and supports a 'knock-off' mechanism of ion permeation through a stepwise-binding process. The multi-ion selectivity filter of our  $\text{Ca}_V\text{Ab}$  model establishes a structural framework for understanding the mechanisms of ion selectivity and conductance by vertebrate  $\text{Ca}_V$  channels.

$\text{Ca}^{2+}$  ions flow through  $\text{Ca}_V$  channels at a rate of  $\sim 10^6$  ions  $\text{s}^{-1}$ , yet  $\text{Na}^+$  conductance is  $>500$ -fold lower<sup>1</sup>. Such high-fidelity, high-throughput  $\text{Ca}_V$  channel performance is important in regulating intracellular processes such as contraction, secretion, neurotransmission and gene expression in many different cell types<sup>2</sup>. Because the extracellular concentration of  $\text{Na}^+$  is 70-fold higher than  $\text{Ca}^{2+}$ , these essential biological functions require  $\text{Ca}_V$  channels to be highly selective for  $\text{Ca}^{2+}$  in preference to  $\text{Na}^+$ , even though  $\text{Ca}^{2+}$  and  $\text{Na}^+$  have nearly identical diameters ( $\sim 2$  Å). Ion selectivity of  $\text{Ca}_V$  channels is proposed to result from high-affinity binding of  $\text{Ca}^{2+}$ , which prevents  $\text{Na}^+$  permeation. Fast  $\text{Ca}^{2+}$  flux through  $\text{Ca}_V$  channels is thought to use a 'knock-off' mechanism in which electrostatic repulsion between  $\text{Ca}^{2+}$  ions within the selectivity filter overcomes tight binding of a single  $\text{Ca}^{2+}$  ion<sup>1,3–8</sup>. Most of these mechanisms require a multi-ion pore, yet extensive mutational analyses of ion selectivity and cation block of vertebrate  $\text{Ca}_V$  channels support a single high-affinity  $\text{Ca}^{2+}$ -binding site<sup>1,9–14</sup>.

$\text{Ca}_V$  channels contain a single ion-selective pore in the centre of four homologous domains<sup>2</sup>. The central pore is lined by the transmembrane segments (S) S5 and S6 and the intervening 'Pore (P)-loop' from each domain in a four-fold pseudosymmetrical arrangement. The four voltage-sensing modules composed of S1–S4 transmembrane helices are symmetrically arranged around the central pore.  $\text{Ca}_V$  channels are members of the voltage-gated ion channel superfamily and are closely related to voltage-gated  $\text{Na}^+$  ( $\text{Na}_V$ ) channels. Three structures of homotetrameric bacterial  $\text{Na}_V$  channels open the way to elucidating the structural basis for ion selectivity and conductance of vertebrate  $\text{Na}_V$  and  $\text{Ca}_V$  channels<sup>15–17</sup>, which probably evolved from the bacterial  $\text{NaChBac}$  family and retained similar structures and functions (Supplementary Fig. 1)<sup>18–20</sup>. Interestingly, mutation of three amino-acid residues in the selectivity filter of  $\text{NaChBac}$  is sufficient to confer  $\text{Ca}^{2+}$  selectivity<sup>21</sup>. We introduced analogous mutations into the bacterial  $\text{Na}_V$  channel  $\text{Na}_V\text{Ab}$  to create  $\text{Ca}_V\text{Ab}$  and carried out electrophysiological and X-ray

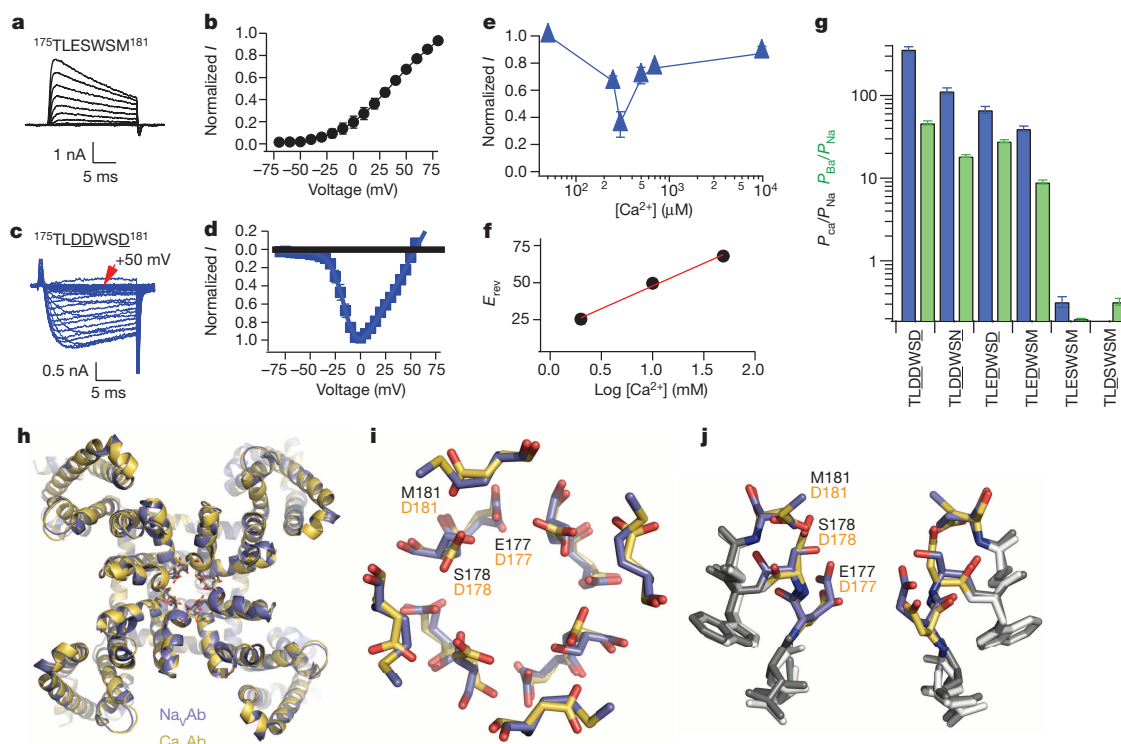
crystallographic analyses to determine the relative permeability of  $\text{Ca}^{2+}$  and define ion-binding sites in the selectivity filter. Our systematic analyses of  $\text{Ca}_V\text{Ab}$  and intermediate derivatives provide structural and mechanistic insights into  $\text{Ca}^{2+}$  binding and ion permeation and suggest a conductance mechanism involving two energetically similar ion-occupancy states with one or two hydrated  $\text{Ca}^{2+}$  ions bound.

## Structure and function of $\text{Ca}_V\text{Ab}$

$\text{Na}_V\text{Ab}$  channels have four identical pore motifs (<sup>175</sup>TLESWSM<sup>181</sup>) that form the ion selectivity filter<sup>15</sup>. The side chains of E177 form a high-field-strength site (Site<sub>HFS</sub>) at the outer end of the filter, whereas two additional potential  $\text{Na}^+$ -coordination sites, a central site (Site<sub>CEN</sub>) and an inner site (Site<sub>IN</sub>), are formed by the backbone carbonyls of L176 and T175 (ref. 15). To create  $\text{Ca}_V\text{Ab}$ , E177, S178 and M181 were substituted with Asp, resulting in a mutant with the pore motif <sup>175</sup>TLDDWSD<sup>181</sup> (underlined letters indicate mutated residues).  $\text{Ca}_V\text{Ab}$  was expressed in *Trichopulsia ni* cells (High5) and analysed by whole-cell voltage clamp to determine its ion selectivity. In contrast to  $\text{Na}_V\text{Ab}$ , which does not conduct extracellular  $\text{Ca}^{2+}$  ions but carries outward  $\text{Na}^+$  current (Fig. 1a, b),  $\text{Ca}_V\text{Ab}$  conducts inward  $\text{Ca}^{2+}$  current in a voltage-dependent manner (Fig. 1c, d). Complete titration curves for  $\text{Ca}^{2+}$  in the presence of  $\text{Ba}^{2+}$  as the balancing divalent cation (see Methods) revealed inhibition of  $\text{Ba}^{2+}$  current by low concentrations of  $\text{Ca}^{2+}$  followed by increases in  $\text{Ca}^{2+}$  current at higher  $\text{Ca}^{2+}$  concentrations (Fig. 1e). These results demonstrate the anomalous mole fraction effect characteristic of vertebrate  $\text{Ca}_V$  channels. Comparable experiments with  $\text{Na}^+$  as the balancing cation were not possible because of the instability of the High5 cells in solutions with low divalent cation concentrations. The reversal potential for  $\text{Ca}^{2+}$  current under bi-ionic conditions closely follows the expectation for a highly  $\text{Ca}^{2+}$ -selective conductance ( $30.6 \pm 2.3$  mV decade<sup>-1</sup>, Fig. 1f and Supplementary Fig. 2), and  $\text{Ca}_V\text{Ab}$  selects  $\text{Ca}^{2+}$  382-fold over  $\text{Na}^+$  under our standard

<sup>1</sup>Department of Pharmacology, University of Washington, Seattle, Washington 98195, USA. <sup>2</sup>Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. <sup>†</sup>Present address: Department of Structural Biology, Genentech Inc., South San Francisco, California 94080, USA.

\*These authors contributed equally to this work.



**Figure 1 | Structure and function of the CaVAb channel.** **a, b,** Outward Na<sup>+</sup> current conducted by NaVAb with 10 mM extracellular Ca<sup>2+</sup> and 140 mM intracellular Na<sup>+</sup>. Holding potential, -100 mV; 20-ms, 10-mV step depolarizations. **c, d,** Voltage-dependent conductance of inward Ca<sup>2+</sup> current by CaVAb under the same conditions. 20-ms, 5-mV step depolarizations. **e,** Biphasic anomalous mole fraction effect of increasing Ca<sup>2+</sup> as indicated, with Ba<sup>2+</sup> as the balancing divalent cation: 10 mM Ba<sup>2+</sup> with 0 to 0.5 mM Ca<sup>2+</sup>, 9.3 mM Ba<sup>2+</sup> with 0.7 mM Ca<sup>2+</sup>, and 0 mM Ba<sup>2+</sup> with 10 mM Ca<sup>2+</sup>

( $n = 4-10$ ). **f,** Reversal potential ( $E_{rev}$ ) versus Ca<sup>2+</sup> concentration. **g,** Relative permeability of CaVAb and its derivatives as measured from bi-ionic reversal potentials.  $P_{Ca}/P_{Na}$ , blue;  $P_{Ba}/P_{Na}$ , green ( $n = 5-22$ ). **h,** Cartoon representation of the overall structure of CaVAb (yellow) superimposed with NaVAb (slate). **i, j,** Top (i) and side (j) views of the superimposed selectivity filters of CaVAb (yellow) and NaVAb (slate) in stick representation. The three original NavAb residues (black) and substituted CavAb residues (orange) are indicated. Errors bars in **b** and **d-g** are  $\pm$  s.e.m.

recording conditions, yielding a range of the permeability ( $P$ ) ratio  $P_{Ca}/P_{Na} > 10,000$ -fold for these constructs (Fig. 1g). Intermediate CaVAb derivatives with single and double Asp substitutions had progressive increases in Ca<sup>2+</sup> selectivity (Fig. 1g and Supplementary Fig. 2), as observed for NaChBac<sup>21</sup>. The <sup>175</sup>TLDDWSN<sup>181</sup> mutant has an Asn residue in place of the final Asp, as observed in one domain of mammalian CaV channels (Supplementary Fig. 1), and it still favours Ca<sup>2+</sup> over Na<sup>+</sup> by more than 100-fold (Fig. 1g).

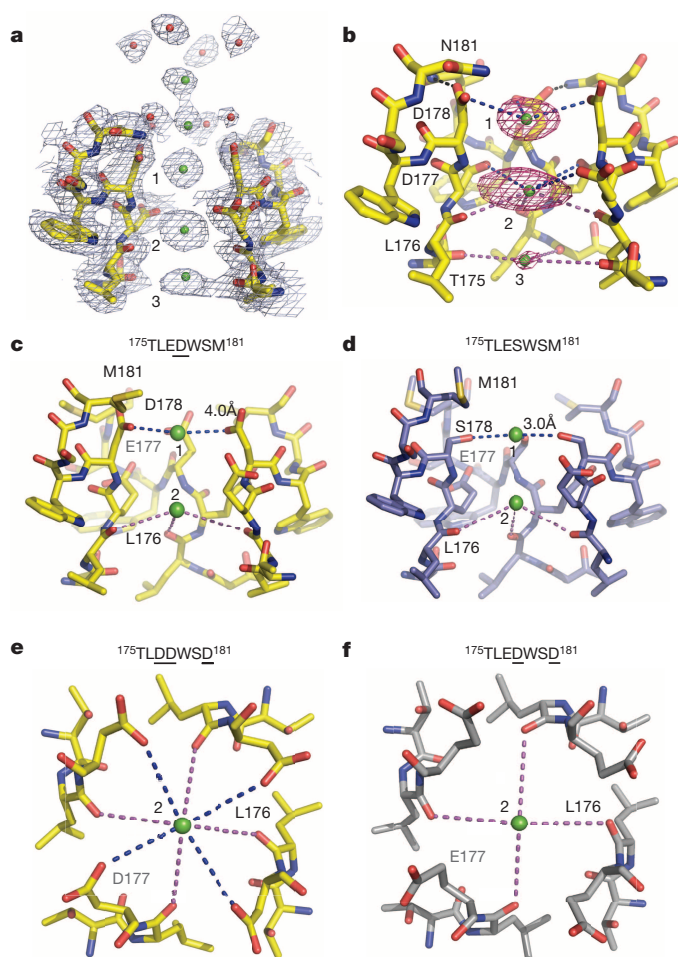
We crystallized and determined the structure of CaVAb and its derivatives by molecular replacement using the NaVAb structure (PDB code 3RVY) as the search template (Supplementary Table 1). The overall structure of CaVAb is very similar to that of NaVAb, with a root mean squared deviation (r.m.s.d.) of 0.4 Å (Fig. 1h). However, the electrostatic potential at the outer entry to the selectivity filter is more negative for CaVAb than for NaVAb (Supplementary Fig. 3). The three negatively charged Asp residues introduced at the selectivity filter of CaVAb create a wide, short, electronegatively lined pore (6 Å diameter, 10 Å length) with no significant alteration in backbone structure with respect to NaVAb (Fig. 1i, j and Supplementary Fig. 4). Thus, the Ca<sup>2+</sup> selectivity of CaVAb is mainly determined by the side chains of the amino acids at the selectivity filter.

### Ca<sup>2+</sup>-binding sites in the permeation pathway

The 3.2 Å resolution structure of the mutant <sup>175</sup>TLDDWSN<sup>181</sup> in the presence of 10 mM Ca<sup>2+</sup> reveals electron densities in the selectivity filter consistent with three Ca<sup>2+</sup> ions aligned on the central axis (Fig. 2a). In the outer vestibule leading to the selectivity filter, there are two additional less-intense on-axis peaks associated with weaker surrounding densities. To confirm the identity of the bound ions, we collected X-ray diffraction data at a wavelength of 1.75 Å and calculated the  $F^+_{Ca} - F^-_{Ca}$

anomalous difference map. Two strong peaks followed by a weaker peak on the intracellular side were found in the selectivity filter along the ion-conduction pathway, verifying three binding sites for Ca<sup>2+</sup> (Fig. 2b). We name these Site 1, Site 2 and Site 3 from the extracellular to the intracellular side.

The Ca<sup>2+</sup> ion at Site 1 is predominantly coordinated by the carboxyl groups of D178 (Site<sub>EX</sub> in NaVAb), which define a plane at the selectivity filter entrance on the extracellular side of the bound Ca<sup>2+</sup> ion (Fig. 2b). The distance between the carboxyl oxygen and Ca<sup>2+</sup> is about 4.0 Å. This distance suggests that the ion binds at this site in a hydrated form because the ionic diameter of Ca<sup>2+</sup> is 2.28 Å, too small to interact with the carboxylate anions directly but appropriate for interaction through bound water molecules. Further into the pore, the four acidic side chains of D177 (Site<sub>HFS</sub> in NaVAb) are located along the wall of the selectivity filter rather than projecting into the lumen, thereby also allowing the binding of a fully hydrated Ca<sup>2+</sup> ion (Fig. 2b). Different from Site 1, this central Ca<sup>2+</sup>-binding site (Site 2) is surrounded by a box of four carboxylate oxygen atoms from D177 above and four backbone carbonyl oxygen atoms from L176 below (Site<sub>CEN</sub> in NaVAb), with oxygen-Ca<sup>2+</sup> distances of 4.5 Å and 4.2 Å, respectively (Fig. 2b). At the intracellular side of the pore, the third Ca<sup>2+</sup>-binding site (Site 3) is composed of one plane of four carbonyls from T175 (Site<sub>IN</sub> in NaVAb), which point inward to the lumen (Fig. 2b). Here the Ca<sup>2+</sup> ion lies nearly on the same plane as T175 carbonyls. The chemical environment of Site 3 hints at a lower affinity, consistent with its role in exit of Ca<sup>2+</sup> from the selectivity filter into the central cavity. Throughout the selectivity filter, the oxygen-Ca<sup>2+</sup> coordination distances are in the range of 4.0–5.0 Å, suggesting that the bound Ca<sup>2+</sup> ion is continuously stabilized in a fully hydrated state when it passes through the pore. We observed diffuse electron density and in favourable cases



**Figure 2 |  $\text{Ca}^{2+}$ -binding sites in and near the selectivity filter of  $\text{NaVAb}$ ,  $\text{CavAb}$  and their derivatives.** **a**, Electron density at the selectivity filter of  $^{175}\text{TLDDWSN}^{181}$  (also see Supplementary Fig. 4). The  $2F_o - F_c$  electron density map (contoured at  $2\sigma$ ) of select residues in the selectivity filter with two diagonally opposed subunits shown in sticks, the  $\text{Ca}^{2+}$  ions along the ion pathway in green spheres and water molecules in red spheres. **b**, Densities at  $\text{Ca}^{2+}$  binding sites 1 and 2 from the anomalous difference Fourier map ( $3\sigma$ ) calculated from the diffraction data of a  $^{175}\text{TLDDWSN}^{181}$  mutant crystal soaked in the presence of 5 mM  $\text{Ca}^{2+}$  and collected at 1.75 Å wavelength. The distances between  $\text{Ca}^{2+}$  and oxygen atoms (dashed lines) are about 4.0 Å at Site 1 (blue lines), 4.4 Å at Site 2 (blue and magenta lines) and 5.0 Å (magenta line) at Site 3. For clarity, the subunit closest to the viewer is not shown. **c, d**, A comparison between  $^{175}\text{TLEDWSM}^{181}$  and  $^{175}\text{TLESWSM}^{181}$  ( $\text{NaVAb}$ ) highlighting the importance of Site 1 for  $\text{Ca}^{2+}$  selectivity. **e, f**, A comparison between  $^{175}\text{TLDDWSN}^{181}$  ( $\text{CavAb}$ ) and  $^{175}\text{TLEDWSN}^{181}$  highlighting the role of Site 2 in fine tuning  $\text{Ca}^{2+}$  selectivity. All structures were determined in the presence of 15 mM  $\text{Ca}^{2+}$ .

discrete water molecules surrounding the bound  $\text{Ca}^{2+}$ , consistent with the presence of an inner shell of bound waters of hydration (Supplementary Fig. 5).

Although the anomalous difference map did not resolve clear peaks at the outer vestibule beyond the selectivity filter, we interpret the two on-axis  $2F_o - F_c$  densities above the three  $\text{Ca}^{2+}$  sites as two additional  $\text{Ca}^{2+}$  ions poised to enter the pore (Fig. 2a). This assignment is supported by the surrounding eight islets of density, which probably represent eight stabilized water molecules. Just as at Site 2 in the selectivity filter, these eight water molecules appear to serve as a square antiprism cage coordinating a hydrated  $\text{Ca}^{2+}$  ion at the centre (Fig. 2a). The second  $\text{Ca}^{2+}$  ion located at the bottom of this cage is  $\sim 4.5$  Å away from the four carboxyl oxygen atoms of D178, suggesting that part of its second hydration shell is replaced by D178 before the ion enters the selectivity filter. The selectivity filter, therefore, appears to select  $\text{Ca}^{2+}$

at its mouth by recognizing the  $\text{Ca}^{2+}$ - $\text{H}_2\text{O}$  hydration complex and conduct  $\text{Ca}^{2+}$  by fitting the  $\text{Ca}^{2+}$ - $\text{H}_2\text{O}$  hydration complex into the pore. Because  $\text{Ca}^{2+}$  is more electropositive than  $\text{Na}^+$ , it should bind more tightly in the ion selectivity filter of  $\text{CavAb}$ , providing a mechanistic basis for the block of  $\text{Na}^+$  permeation by  $\text{Ca}^{2+}$  at low  $\text{Ca}^{2+}$  concentration and preferential permeation of  $\text{Ca}^{2+}$  at higher  $\text{Ca}^{2+}$  concentration (see Discussion).

### Functional roles of key selectivity filter residues

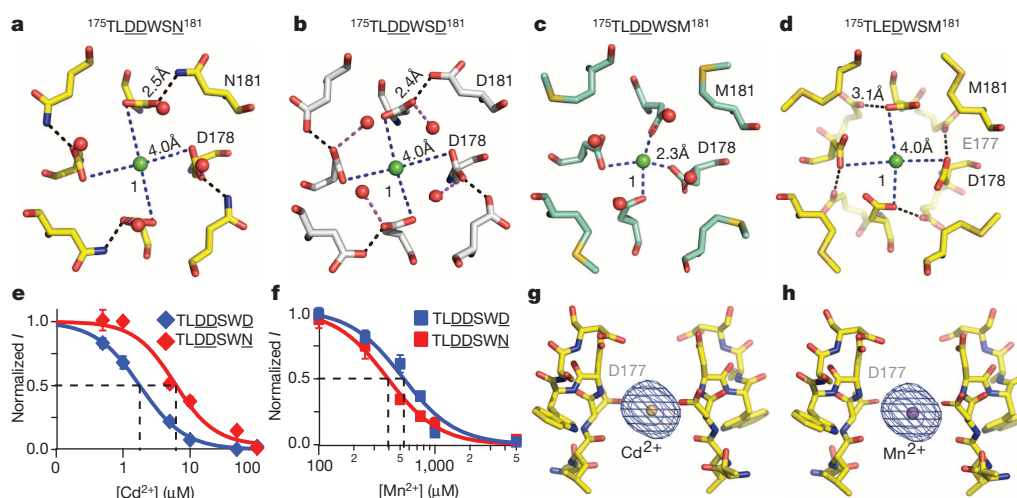
Measurements of bi-ionic reversal potentials revealed that the relative permeability of different  $\text{CavAb}$  intermediate constructs for  $\text{Ca}^{2+}$  follows the order of  $\text{CavAb}$  ( $^{175}\text{TLDDWSN}^{181}$ )  $>$   $^{175}\text{TLDDWSM}^{181}$   $>$   $^{175}\text{TLEDWSM}^{181}$   $>$   $^{175}\text{TLESWSM}^{181}$  (Fig. 1g and Supplementary Fig. 2). A comparison of the  $\text{Ca}^{2+}$  selectivity ratios between  $^{175}\text{TLEDWSM}^{181}$  and  $^{175}\text{TLESWSM}^{181}$  ( $\text{NaVAb}$ ) shows that substitution of S178 with Asp is sufficient to convert the selectivity from  $\text{Na}^+$  to  $\text{Ca}^{2+}$  with  $>100$ -fold change in  $P_{\text{Ca}}:P_{\text{Na}}$  (Fig. 1g). Placement of the Asp carboxyl side chain at this position allows for the formation of the first hydrated  $\text{Ca}^{2+}$ -binding site in the selectivity filter (Fig. 2c and Supplementary Fig. 6). By contrast, S178 in  $\text{NaVAb}$  binds  $\text{Ca}^{2+}$  directly by displacing its hydration shell, which blocks conductance of both  $\text{Na}^+$  and  $\text{Ca}^{2+}$  (Fig. 2d). Therefore, formation of Site 1 for binding hydrated  $\text{Ca}^{2+}$  is both necessary and sufficient for conferring  $\text{Ca}^{2+}$  selectivity over  $\text{Na}^+$  to  $\text{NaVAb}$ .

The  $\text{Ca}^{2+}$  selectivity ratio of  $\text{CavAb}$  ( $^{175}\text{TLDDWSN}^{181}$ ) is 5.5-fold higher than  $^{175}\text{TLEDWSN}^{181}$  (Fig. 1g). This functional difference reflects a role of Site 2 in adjusting  $\text{Ca}^{2+}$  selectivity. Different from the side chains of D177 in  $\text{CavAb}$  ( $^{175}\text{TLDDWSN}^{181}$ ), which interact with the  $\text{Ca}^{2+}$  ion (Fig. 2e), the carboxyl group of E177 in  $^{175}\text{TLEDWSN}^{181}$  swings away from the selectivity filter and forms a hydrogen bond with D181 and the main-chain nitrogen atoms of S180 (Fig. 2f and Supplementary Fig. 7). Site 2 in  $^{175}\text{TLEDWSN}^{181}$ , therefore, is exclusively formed by the four carbonyl oxygen atoms of L176, which conceivably leads to a lower  $\text{Ca}^{2+}$ -binding affinity and a decreased  $\text{Ca}^{2+}$  selectivity. This comparison highlights both the importance of Site 2 in supporting high  $\text{Ca}^{2+}$  selectivity and the critical role of the backbone carbonyl groups of L176 in constructing this ion-binding site.

Distinct from D177 and D178, the N181 residue of  $^{175}\text{TLDDWSN}^{181}$  lies outside of the ion-conducting pore and is not directly involved in  $\text{Ca}^{2+}$  ion coordination. In close proximity to the carboxyl groups of D178, which form a ring that lines the perimeter of the pore entryway, the side chain of N181 embraces the perimeter of the D178 ring by donating a hydrogen bond to its side-chain carboxyls (Fig. 3a). Such a structural arrangement is also found in  $\text{CavAb}$  ( $^{175}\text{TLDDWSN}^{181}$ ) (Fig. 3b), although the more electronegative environment created by the extra negatively charged residue, D181, probably attracts  $\text{Ca}^{2+}$  more strongly and confers a 4- to 5-fold higher degree of  $\text{Ca}^{2+}$  selectivity to  $\text{CavAb}$  ( $^{175}\text{TLDDWSN}^{181}$ ) in comparison to  $^{175}\text{TLDDWSM}^{181}$  (Fig. 1g and Supplementary Fig. 3).

$^{175}\text{TLDDWSM}^{181}$ , which has the hydrophobic residue M181 packed next to the D178 ring, is the only  $\text{CavAb}$  intermediate that does not conduct  $\text{Ca}^{2+}$  (Supplementary Fig. 2). The crystal structure of this mutant reveals a blocking  $\text{Ca}^{2+}$  ion tightly bound at Site 1 in a dehydrated state with an oxygen-ion distance of 2.3 Å (Fig. 3c). Superposition analysis shows few structural differences between  $^{175}\text{TLDDWSM}^{181}$  and  $^{175}\text{TLDDWSN}^{181}$ , except for the side chain of D178, which is fixed by N181 in  $^{175}\text{TLDDWSN}^{181}$  but unconstrained in  $^{175}\text{TLDDWSM}^{181}$  (Fig. 3a, c). This comparison indicates that N181 in  $^{175}\text{TLDDWSN}^{181}$  and D181 in  $\text{CavAb}$  have critical roles in engaging D178 and allowing the reversible binding of the  $\text{Ca}^{2+}$ - $\text{H}_2\text{O}$  hydration complex for active  $\text{Ca}^{2+}$  conductance. Although the subtle difference in  $\text{Ca}^{2+}$  selectivity between  $^{175}\text{TLEDWSN}^{181}$  and  $^{175}\text{TLEDWSM}^{181}$  seems to argue against this conclusion (Fig. 1g), E177 in  $^{175}\text{TLEDWSM}^{181}$  actually has a structural role equivalent to that of N181 in  $^{175}\text{TLDDWSN}^{181}$  — by pointing away from the selectivity filter lumen, E177 forms a carboxylate-carboxylate





**Figure 3 | Ion binding and block of  $\text{Ca}_v\text{Ab}$  and its derivatives.** **a, b,** Top view of Site 1 with a hydrated  $\text{Ca}^{2+}$  ion coordinated by D178 with the help of N181 and D181 in  $^{175}\text{TLDDWSN}^{181}$  and  $^{175}\text{TLDDWSD}^{181}$  ( $\text{Ca}_v\text{Ab}$ ), respectively. **c,** Binding of a dehydrated  $\text{Ca}^{2+}$  ion at Site 1 in the nonconductive  $^{175}\text{TLDDWSM}^{181}$  mutant. **d,** Coordination of a hydrated  $\text{Ca}^{2+}$  ion at Site 1 of the  $^{175}\text{TLEDWSM}^{181}$  mutant. Despite the absence of a polar residue at amino acid 181, E177 in  $^{175}\text{TLEDWSM}^{181}$  is able to hold D178 in place to allow the binding of a hydrated  $\text{Ca}^{2+}$  ion. **e, f,** Block of  $\text{Ca}^{2+}$  conductance by the

indicated concentrations of  $\text{Cd}^{2+}$  and  $\text{Mn}^{2+}$ .  $^{175}\text{TLDDWSD}^{181}$ :  $\text{IC}_{50}(\text{Cd}^{2+})$ ,  $1.7 \pm 0.04 \mu\text{M}$ ;  $\text{IC}_{50}(\text{Mn}^{2+})$ ,  $526 \pm 22 \mu\text{M}$ .  $^{175}\text{TLDDWSN}^{181}$ :  $\text{IC}_{50}(\text{Cd}^{2+})$ ,  $5.9 \pm 0.4 \mu\text{M}$ ;  $\text{IC}_{50}(\text{Mn}^{2+})$ ,  $388 \pm 7 \mu\text{M}$ . Error bars are  $\pm$  s.e.m. **g, h,** Side view of the  $\text{Cd}^{2+}$ - and  $\text{Mn}^{2+}$ -binding sites in the selectivity filter of  $\text{Ca}_v\text{Ab}$ . The anomalous difference Fourier map densities (blue mesh, contoured at  $5\sigma$ ) of the bound blocking ions are calculated using diffraction data collected at  $1.75 \text{ \AA}$  wavelength. For clarity, the residues forming the selectivity filter from the two subunits in front of and behind the plane of the drawing were removed.

pair with D178 and holds it in a conduction-competent position (Fig. 3d and Supplementary Fig. 8).

### Block of $\text{Na}_v\text{Ab}$ and $\text{Ca}_v\text{Ab}$ channels by divalent cations

$\text{Cd}^{2+}$ ,  $\text{Mn}^{2+}$  and other inorganic cations are effective blockers of  $\text{Ca}_v\text{Ab}$  channels<sup>1</sup>. Block of  $\text{Ca}^{2+}$  conductance of  $\text{Ca}_v\text{Ab}$  by  $\text{Cd}^{2+}$  and  $\text{Mn}^{2+}$  gives  $K_i$  values of  $1.78 \mu\text{M}$  for  $\text{Cd}^{2+}$  and  $526 \mu\text{M}$  for  $\text{Mn}^{2+}$  (Fig. 3e, f, blue).  $\text{Cd}^{2+}$  has a lower affinity and  $\text{Mn}^{2+}$  has a higher affinity for block of  $^{175}\text{TLDDWSN}^{181}$  (Fig. 3e, f, red). Crystals with bound  $\text{Cd}^{2+}$  and  $\text{Mn}^{2+}$  were obtained by soaking  $\text{Ca}_v\text{Ab}$  crystals in a cryo-solution containing these heavy metal ions, and the anomalous difference map was calculated from a data set collected at  $1.75 \text{ \AA}$  wavelength. The structures show that both  $\text{Cd}^{2+}$  and  $\text{Mn}^{2+}$  bind in the selectivity filter at the central site (Site 2), which is coordinated by the side chains of the four D177 residues and the carbonyl groups of L176 (Fig. 3g, h). Locked at this site, these blocking ions would inhibit the  $\text{Ca}^{2+}$  current by competitively binding to the high-affinity site required for  $\text{Ca}^{2+}$  permeation. Another important common feature of the two blocking complexes of  $\text{Ca}_v\text{Ab}$  is the block of permeation by binding of a single divalent cation within the selectivity filter, which supports the hypothesis that at least two divalent-cation-binding sites must be located close enough to induce repulsive interactions and allow divalent cation conductance by a knock-off mechanism. Because they are smaller than  $\text{Ca}^{2+}$ , the bound  $\text{Cd}^{2+}$  ( $d = 2.18 \text{ \AA}$ ) and  $\text{Mn}^{2+}$  ( $d = 1.94 \text{ \AA}$ ) must interact with the selectivity filter through bound waters of hydration, and electron density consistent with bound waters of hydration is observed in our structures (Supplementary Fig. 5).

### Ion binding at the $\text{Ca}^{2+}$ selectivity filter

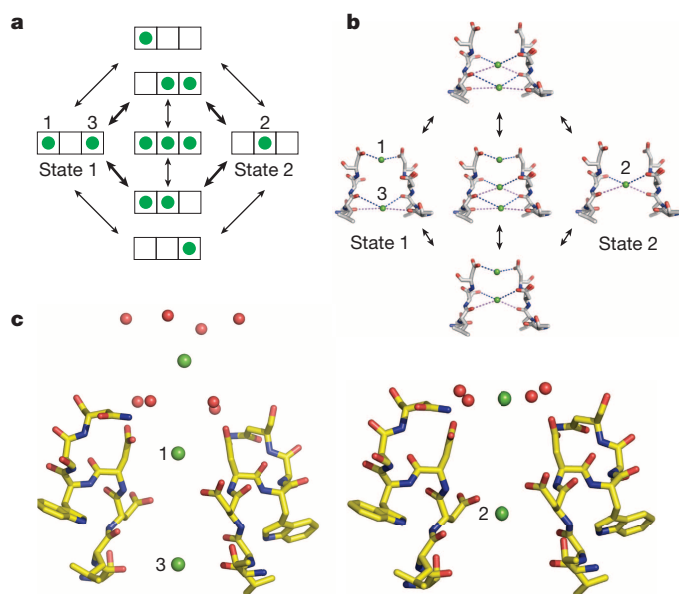
To assess the properties of the three  $\text{Ca}^{2+}$ -binding sites in the selectivity filter of  $^{175}\text{TLDDWSN}^{181}$ , we titrated the concentration of  $\text{Ca}^{2+}$  in the cryo-solution and calculated the anomalous difference maps. At low  $\text{Ca}^{2+}$  concentration, two strong peaks of approximately equal intensity are found at Site 1 and Site 2 (Supplementary Fig. 9). As the  $\text{Ca}^{2+}$  concentration is raised, the electron density of Site 2 is substantially enhanced, but the peak intensity is reduced at Site 1 and remains low at Site 3 (Supplementary Fig. 9). These results suggest that the central site has the highest affinity, whereas Site 3 is the weakest. It is probable that this titration pattern reflects independent binding of  $\text{Ca}^{2+}$  to Sites 1, 2

and 3 located in different individual molecules of  $\text{Ca}_v\text{Ab}$  at low  $\text{Ca}^{2+}$  concentration, whereas increasing concentrations of  $\text{Ca}^{2+}$  saturate Site 2 in most or all individual  $\text{Ca}_v\text{Ab}$  molecules and reduce or eliminate binding at Sites 1 and 3 by repulsion. Importantly, the two flanking sites have lower affinity than the central site, as proposed in the 'stepwise binding model' of  $\text{Ca}_v$  channel permeation<sup>7</sup>. In this model, the presence of flanking sites of intermediate affinity facilitates the movement of  $\text{Ca}^{2+}$  into and out of a central high-affinity site, which can result in high ion conductance, even in the limiting case where there is no repulsion between bound ions.

Consistent with high binding affinity,  $\text{Ca}^{2+}$  binds at Site 2 with its first hydration shell coordinated with eight oxygen atoms from the channel (Fig. 2b and Supplementary Fig. 5). By contrast,  $\text{Ca}^{2+}$  at site 1 is mainly stabilized by one plane of four carboxyl groups from D178. The distance between the  $\text{Ca}^{2+}$  ion at Site 1 and the carboxyl group of D177 at Site 2 is about  $5.5\text{--}6 \text{ \AA}$ . As the  $\text{Ca}^{2+}$  ion moves inward, this distance will be reduced enough for D177 to form a stable coordination with the moving  $\text{Ca}^{2+}$  ion. This spatial configuration suggests that the two sites are separated by a low energy barrier. The differences of negative charge between D178 and the carbonyls of T175 and the differences in the geometry of their interactions with  $\text{Ca}^{2+}$  provide a plausible explanation for the higher  $\text{Ca}^{2+}$ -binding affinity at Site 1 than Site 3.

### Ion permeation mechanism

The three  $\text{Ca}^{2+}$ -binding sites in the selectivity filter of  $^{175}\text{TLDDWSN}^{181}$  are separated by a distance of about  $4.5 \text{ \AA}$ , which would result in substantial electrostatic repulsive interactions between bound ions. As in the case of the KcsA channel<sup>22</sup>, it is energetically unfavourable for  $\text{Ca}^{2+}$  ions to occupy adjacent sites simultaneously. This leads directly to our hypothesis of two interchangeable functional states of the selectivity filter in the crystal structure (Fig. 4a, b). In State 1,  $\text{Ca}^{2+}$  ions occupy Site 1 and Site 3. In State 2, a single  $\text{Ca}^{2+}$  ion occupies Site 2. These two states might be further coupled with one of the two  $\text{Ca}^{2+}$  ions at the outer vestibule ready to enter the pore (Fig. 4c). The transition between these two states occurs either when  $\text{Ca}^{2+}$  jumps from Site 1 or 3 to Site 2 or a third ion enters on one side of the filter, causing an ion to move into Site 2. It is probable that our crystal structures reflect a mixed population of  $\text{Ca}_v\text{Ab}$  molecules in which only Site 2 is occupied by  $\text{Ca}^{2+}$  plus  $\text{Ca}_v\text{Ab}$  molecules in which Site 1 and/or Site 3 are occupied. Because of



**Figure 4 | Catalytic cycle for Ca<sup>2+</sup> conductance by Ca<sub>V</sub>Ab.** **a**, An ionic occupancy state diagram of Ca<sub>V</sub>Ab showing two proposed low energy states and the potential transitions that connect them. Each state of the selectivity filter is represented by a three-box rectangle with Sites 1–3 going from left to right. Green circles represent Ca<sup>2+</sup> ions. Note that transitions in the inner circle potentially lead to ion repulsion, which might facilitate conduction. These transitions in the inner circle are more probable than those in the outer circle, as denoted by the bold arrows. **b**, The structural basis of the ionic occupancy states depicted in the inner circle of the state diagram shown on the left. The clockwise cycle represents a path for inward flux of Ca<sup>2+</sup> ions through the selectivity filter. **c**, Coupling of extracellular Ca<sup>2+</sup>-binding sites and the three sites within the selectivity filter in the two proposed ionic occupancy states. When two Ca<sup>2+</sup> ions bind to position 1 and 3 in the filter, the entryway Ca<sup>2+</sup> ion is placed furthest from the pore (left). When one Ca<sup>2+</sup> ion binds to position 2 within the filter, the ion outside the filter is pulled closer to the pore (right).

the high concentration of Ca<sup>2+</sup> in the extracellular solution, Ca<sup>2+</sup> will prefer to enter Site 1 and the weak binding of Ca<sup>2+</sup> to Site 3 will force loss of Ca<sup>2+</sup> into the low Ca<sup>2+</sup> concentration in the cytosol. This generates a unidirectional flux of Ca<sup>2+</sup> into the cell (Fig. 4c). The three-ion-occupied state would be manifest only when the external Ca<sup>2+</sup> concentration is increased enough that the flux reaches a limiting value<sup>3</sup>. The presence of the lower-affinity Site 3 flanking the central cavity would further accelerate the flux of ions by allowing stepwise binding with relatively low chemical potential energy barriers<sup>7</sup>. The combination of ionic repulsion between Ca<sup>2+</sup> ions bound at these sites and their stepwise change in binding affinity work together to allow rapid conduction in spite of the intrinsic high affinity for Ca<sup>2+</sup> binding.

## Discussion

The mechanism underlying the dramatic difference in selectivity for Ca<sup>2+</sup> over Na<sup>+</sup> in Ca<sub>V</sub>Ab versus Na<sub>V</sub>Ab is different from the mechanisms responsible for selectivity of K<sup>+</sup> over Na<sup>+</sup> and for Ca<sup>2+</sup> block revealed by high-resolution structural studies of the NaK channel<sup>23–25</sup>. In the NaK channel, K<sup>+</sup> conductance is favoured by the presence of four binding sites formed by backbone carbonyls, rather than two or three, and structural changes in amino acid residues outside the ion selectivity filter fine-tune the electronegativity of backbone carbonyls in the selectivity filter and thereby determine the affinity for block by Ca<sup>2+</sup> at the extracellular mouth of the pore. This difference in the underlying mechanism for control of ion selectivity reflects the fundamental difference in ion permeation in Na<sub>V</sub>Ab and Ca<sub>V</sub>Ab versus K<sup>+</sup> channels. In Na<sub>V</sub>Ab and Ca<sub>V</sub>Ab, permeant ions interact with both amino acid side chains and backbone carbonyls in the ion selectivity filter primarily through waters of hydration, whereas K<sup>+</sup> channels

select their permeant ions through direct interaction of the dehydrated ions with backbone carbonyls.

Our results reveal an unexpected structural basis for Ca<sup>2+</sup> selectivity and conductance in Ca<sub>V</sub>Ab channels, in which most or all interactions of Ca<sup>2+</sup> with the pore are made through its inner shell of waters of hydration. A set of three Ca<sup>2+</sup> binding sites cooperate in a knock-off mechanism in which the selectivity filter oscillates primarily between two states with either one hydrated Ca<sup>2+</sup> bound at the central site or two hydrated Ca<sup>2+</sup> ions bound at the distal sites. The high-affinity binding of Ca<sup>2+</sup> to Sites 1 and 2 ensures that Na<sup>+</sup> and other monovalent cations cannot permeate, while the high Ca<sup>2+</sup> concentration in the extracellular solution enables unidirectional flux by driving rapid occupancy of Site 1. The ionic repulsion between Ca<sup>2+</sup> ions bound at these sites and their stepwise change in binding affinity work together to allow rapid conduction in spite of the intrinsic high affinity for Ca<sup>2+</sup> binding. Although our resolution does not allow us to see all of the waters of hydration that are implied by our structure, we do observe electron density surrounding bound Ca<sup>2+</sup> ions at Sites 1, 2 and 3 that we believe represents the inner shell of waters of hydration (Supplementary Figure 5). This electron density is blurred, as if there is a diversity of arrangements of the bound water molecules in individual Ca<sub>V</sub>Ab molecules in our crystals because their hydrogen-bonding requirements can be accommodated in multiple ways between the bound cations and their coordinating carboxyl and carbonyl oxygens that comprise Sites 1, 2 and 3. In our most favourable structure (Supplementary Fig. 5g, h), four discrete water molecules are observed at Site 3. Altogether, we believe that these images provide direct support for the conclusion that bound Ca<sup>2+</sup> ions are surrounded by an inner shell of waters of hydration that are dynamic and can easily exchange local hydrogen-binding partners. This is a unique ion conduction mechanism, which allows high-affinity interaction of hydrated Ca<sup>2+</sup> ions while mediating their rapid movement from the extracellular vestibule, through the three ion coordination sites of the selectivity filter, through the central cavity, and finally into the cytosol.

Biophysical modelling of Ca<sup>2+</sup> permeation in vertebrate Ca<sub>V</sub> channels has led to multiple proposed mechanisms, most of which involve two or more Ca<sup>2+</sup>-binding sites, yet only a single high-affinity site that is required for both permeation and Ca<sup>2+</sup> block was identified by mutagenesis and physiological analyses<sup>1</sup>. Our results with Ca<sub>V</sub>Ab channels resolve this apparent discrepancy by showing that multiple Ca<sup>2+</sup>-binding sites are necessary for permeation, but only Site 2 binds divalent cations with sufficient affinity for block. Ca<sup>2+</sup> is conducted as a hydrated cation (Supplementary Fig. 5), consistent with the large estimated functional diameter of vertebrate Ca<sub>V</sub> channels of 6 Å (ref. 26). Detailed structure–function studies of vertebrate Ca<sub>V</sub> channels show that mutations of the four residues equivalent to E177 have distinct effects on Ca<sup>2+</sup> conductance and block, implying that domain-specific interactions with Ca<sup>2+</sup> have evolved in vertebrate four-domain Ca<sub>V</sub> channels<sup>10,11,27–29</sup>. Vertebrate Ca<sub>V</sub> channels might share similar molecular mechanisms for Ca<sup>2+</sup> permeation and selectivity despite their pseudosymmetrical four-domain configuration.

**Note added in proof:** Crystal structures of isolated pore domains of other bacterial Na<sub>V</sub> channels reveal an open pore conformation for Na<sub>V</sub>Ms (ref. 30) and a binding site for blocking Ca<sup>2+</sup> ions in Na<sub>V</sub>Ae (ref. 31), which is formed primarily by the equivalent of Ser 178 in Na<sub>V</sub>Ab.

## METHODS SUMMARY

Ca<sub>V</sub>Ab and its derivative constructs were expressed in *Trichopulsia ni* insect cells and purified using anti-Flag resin and size-exclusion chromatography, reconstituted into DMPC:CHAPSO bicelles, and crystallized over an ammonium sulphate solution containing 0.1 M Na-citrate, pH 4.75. The anomalous data sets were collected at 1.75 Å wavelength with crystals soaked in a stabilizing solution containing various concentrations of cation ions. Electrophysiological experiments were performed in *T. ni* cells using standard protocols.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 17 May; accepted 9 October 2013.**

**Published online 24 November 2013.**

- Sather, W. A. & McCleskey, E. W. Permeation and selectivity in calcium channels. *Annu. Rev. Physiol.* **65**, 133–159 (2003).
- Catterall, W. A. Voltage-gated calcium channels. *Cold Spring Harb. Perspect. Biol.* **3**, a003947 (2011).
- Almers, W. & McCleskey, E. W. Non-selective conductance in calcium channels of frog muscle: calcium selectivity in a single-file pore. *J. Physiol. (Lond.)* **353**, 585–608 (1984).
- Almers, W., McCleskey, E. W. & Palade, P. T. A non-selective cation conductance in frog muscle membrane blocked by micromolar external calcium ions. *J. Physiol. (Lond.)* **353**, 565–583 (1984).
- Hess, P. & Tsien, R. W. Mechanism of ion permeation through calcium channels. *Nature* **309**, 453–456 (1984).
- Armstrong, C. M. & Neyton, J. Ion permeation through calcium channels. A one-site model. *Ann. NY Acad. Sci.* **635**, 18–25 (1991).
- Dang, T. X. & McCleskey, E. W. Ion channel selectivity through stepwise changes in binding affinity. *J. Gen. Physiol.* **111**, 185–193 (1998).
- Lopin, K. V., Obejero-Paz, C. A. & Jones, S. W. Evaluation of a two-site, three-barrier model for permeation in Ca<sub>v</sub>3.1 ( $\alpha$ 1G) T-type calcium channels: Ca<sup>2+</sup>, Ba<sup>2+</sup>, Mg<sup>2+</sup>, and Na<sup>+</sup>. *J. Membr. Biol.* **235**, 131–143 (2010).
- Heinemann, S. H., Terlau, H., Stuhmer, W., Imoto, K. & Numa, S. Calcium channel characteristics conferred on the sodium channel by single mutations. *Nature* **356**, 441–443 (1992).
- Ellinor, P. T., Yang, J., Sather, W. A., Zhang, J. F. & Tsien, R. W. Ca<sup>2+</sup> channel selectivity at a single locus for high-affinity Ca<sup>2+</sup> interactions. *Neuron* **15**, 1121–1132 (1995).
- Yang, J., Ellinor, P. T., Sather, W. A., Zhang, J. F. & Tsien, R. W. Molecular determinants of Ca<sup>2+</sup> selectivity and ion permeation in L-type Ca<sup>2+</sup> channels. *Nature* **366**, 158–161 (1993).
- Kim, M. S., Morii, T., Sun, L. X., Imoto, K. & Mori, Y. Structural determinants of ion selectivity in brain calcium channel. *FEBS Lett.* **318**, 145–148 (1993).
- Cibulsky, S. M. & Sather, W. A. The EEEE locus is the sole high-affinity Ca<sup>2+</sup> binding structure in the pore of a voltage-gated Ca<sup>2+</sup> channel: block by Ca<sup>2+</sup> entering from the intracellular pore entrance. *J. Gen. Physiol.* **116**, 349–362 (2000).
- Cloues, R. K., Cibulsky, S. M. & Sather, W. A. Ion interactions in the high-affinity binding locus of a voltage-gated Ca<sup>2+</sup> channel. *J. Gen. Physiol.* **116**, 569–586 (2000).
- Payandeh, J., Scheuer, T., Zheng, N. & Catterall, W. A. The crystal structure of a voltage-gated sodium channel. *Nature* **475**, 353–358 (2011).
- Payandeh, J., Gamal El-Din, T. M., Scheuer, T., Zheng, N. & Catterall, W. A. Crystal structure of a voltage-gated sodium channel in two potentially inactivated states. *Nature* **486**, 135–139 (2012).
- Zhang, X. *et al.* Crystal structure of an orthologue of the NaChBac voltage-gated sodium channel. *Nature* **486**, 130–134 (2012).
- Ren, D. *et al.* A prokaryotic voltage-gated sodium channel. *Science* **294**, 2372–2375 (2001).
- Yu, F. H. & Catterall, W. A. The VGL-chanome: a protein superfamily specialized for electrical signaling and ionic homeostasis. *Sci. STKE* **2004**, re15 (2004).
- Koishi, R. *et al.* A superfamily of voltage-gated sodium channels in bacteria. *J. Biol. Chem.* **279**, 9532–9538 (2004).
- Yue, L., Navarro, B., Ren, D., Ramos, A. & Clapham, D. E. The cation selectivity filter of the bacterial sodium channel, NaChBac. *J. Gen. Physiol.* **120**, 845–853 (2002).
- Morais-Cabral, J. H., Zhou, Y. & MacKinnon, R. Energetic optimization of ion conduction rate by the K<sup>+</sup> selectivity filter. *Nature* **414**, 37–42 (2001).
- Alam, A. & Jiang, Y. Structural analysis of ion selectivity in the NaK channel. *Nature Struct. Mol. Biol.* **16**, 35–41 (2009).
- Alam, A., Shi, N. & Jiang, Y. Structural insight into Ca<sup>2+</sup> specificity in tetrameric cation channels. *Proc. Natl Acad. Sci. USA* **104**, 15334–15339 (2007).
- Derebe, M. G. *et al.* Tuning the ion selectivity of tetrameric cation channels by changing the number of ion binding sites. *Proc. Natl Acad. Sci. USA* **108**, 598–602 (2011).
- McCleskey, E. W. & Almers, W. The Ca channel in skeletal muscle is a large pore. *Proc. Natl Acad. Sci. USA* **82**, 7149–7153 (1985).
- Chen, X. H. & Tsien, R. W. Aspartate substitutions establish the concerted action of P-region glutamates in repeats I and III in forming the protonation site of L-type Ca<sup>2+</sup> channels. *J. Biol. Chem.* **272**, 30002–30008 (1997).
- Cibulsky, S. M. & Sather, W. A. Control of ion conduction in L-type Ca<sup>2+</sup> channels by the concerted action of S5–6 regions. *Biophys. J.* **84**, 1709–1719 (2003).
- Williamson, A. V. & Sather, W. A. Nonglutamate pore residues in ion selection and conduction in voltage-gated Ca<sup>2+</sup> channels. *Biophys. J.* **77**, 2575–2589 (1999).
- McCusker, E. C. *et al.* Structure of a bacterial voltage-gated sodium channel pore reveals mechanisms of opening and closing. *Nature Commun.* **3**, 1102 (2012).
- Shaya, D. *et al.* Structure of a prokaryotic sodium channel pore reveals essential gating elements and an outer ion binding site common to eukaryotic channels. *J. Mol. Biol.* <http://dx.doi.org/10.1016/j.jmb.2013.10.010> (published online 10 October 2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We are grateful to the beamline staff at the Advanced Light Source (BL8.2.1 and BL8.2.2) for their assistance during data collection. Research reported in this publication was supported by the National Institute of Neurological Disorders and Stroke (NINDS) of the National Institutes of Health (NIH) under award number R01NS015751 (W.A.C.), the National Heart, Lung, and Blood Institute (NHLBI) of the NIH under award number R01HL112808 (W.A.C. and N.Z.) and a National Research Service Award from training grant T32GM008268 (T.M.H.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This work was also supported by the Howard Hughes Medical Institute (N.Z.).

**Author Contributions** L.T., T.M.G.E.-D., J.P., T.S., N.Z. and W.A.C. designed the experiments. J.P. initiated the experimental work. L.T. conducted the protein purification, crystallization and diffraction experiments. L.T., J.P. and N.Z. determined and analysed the structures of the apo and cation-bound forms of Ca<sub>v</sub>Ab and the intermediate Ca<sub>v</sub>Ab constructs. T.M.G.E.-D. and T.S. performed physiological studies of Ca<sub>v</sub>Ab and related constructs. G.Q.M. and T.M.H. made the constructs and performed the preliminary data collection. All authors interpreted the structures in light of the physiological data. L.T., N.Z. and W.A.C. wrote the manuscript with input from all co-authors. W.A.C. and N.Z. are co-senior authors.

**Author Information** Coordinates and structure factors have been deposited in the Protein Data Bank under accession codes: 4MS2 (TLDDWSN, 15 mM Ca<sup>2+</sup>), 4MTF (TLDDWSN, 0.5 mM Ca<sup>2+</sup>), 4MTG (TLDDWSN, 2.5 mM Ca<sup>2+</sup>), 4MTO (TLDDWSN, 5 mM Ca<sup>2+</sup>), 4MVM (TLDDWSN, 10 mM Ca<sup>2+</sup>), 4MVO (TLDDWSN, 15 mM Ca<sup>2+</sup>), 4MVQ (TLDDWSN, 15 mM Ca<sup>2+</sup>), 4MVR (TLDDWSN, 100 mM Mn<sup>2+</sup>), 4MVS (TLDDWSN, 100 mM Cd<sup>2+</sup>), 4MVZ (TLDDWSN, 15 mM Ca<sup>2+</sup>), 4MW3 (TLDDWSN, 15 mM Ca<sup>2+</sup>), 4MVU (TLDDWSN, 15 mM Ca<sup>2+</sup>), 4MW8 (NavAb, 15 mM Ca<sup>2+</sup>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.Z. (nzheng@uw.edu) or W.A.C. (wcatt@uw.edu).



## METHODS

**Protein expression and purification.** The pFastBac-Flag-NavAb(I217C) that was used as the genetic background for CavAb constructs was described previously<sup>15,16</sup>. CavAb and its derivatives—<sup>175</sup>TLDDWSN<sup>181</sup>, <sup>175</sup>TLEDWSD<sup>181</sup>, <sup>175</sup>TLEDWSM<sup>181</sup> and <sup>175</sup>TLDSSWM<sup>181</sup>—were generated via site-directed mutagenesis using Quick-Change (Stratagene). Recombinant baculovirus were produced using the Bac-to-Bac system (Invitrogen), and *T. ni* insect cells were infected for large-scale protein purification. Cells were collected 72 h after infection and re-suspended in buffer A (50 mM Tris-HCl, pH 8.0, 200 mM NaCl) supplemented with protease inhibitors and DNase. After sonication, digitonin (EMD Biosciences) was added to 1% and solubilization was carried out for 1–2 h at 4 °C. Clarified supernatant was then incubated with anti-Flag M2-agarose resin (Sigma) for 1–2 h at 4 °C with gentle mixing. Flag-resin was washed with ten column volumes of buffer B (buffer A supplemented with 0.12% digitonin) and eluted with buffer B supplemented with 0.1 mg ml<sup>-1</sup> Flag peptide. The eluent was concentrated and then passed over a Superdex 200 column (GE Healthcare) in 10 mM Tris-HCl, pH 8.0, 100 mM NaCl and 0.12% digitonin. The peak fractions were concentrated using a Vivaspin 30k centrifugal device.

**Crystallization and data collection.** CavAb and its derivatives were concentrated to ~20 mg ml<sup>-1</sup> and reconstituted into DMPC:CHAPSO (Anatrace) bicelles according to standard protocols<sup>32,33</sup>. The protein–bicelle preparation and a well solution containing 1.8–2.0 M ammonium sulphate, 100 mM Na-citrate, pH 5.0, was mixed with a 1:1 ratio and set up in a hanging-drop vapour-diffusion format. The Ca<sup>2+</sup>-derivative crystals were obtained by soaking CavAb and other mutant crystals in a cryo-protection solution (0.1 M Na-acetate, pH 5.0, 26% glucose and 2.0 M ammonium sulphate) containing the indicated concentrations of Ca<sup>2+</sup> for 40–60 min at 4 °C. The Cd<sup>2+</sup> and Mn<sup>2+</sup> derivatives were obtained by soaking CavAb in the presence of 100 mM Cd<sup>2+</sup> or Mn<sup>2+</sup>, respectively. Crystals were then plunged into liquid nitrogen and maintained at 100 K during all data collection procedures.

All anomalous diffraction data sets were collected at 1.75 Å with the same synchrotron radiation source (Advanced Light Source, BL8.2.1). To optimize the anomalous signal, the data sets were collected by using the ‘inverse beam strategy’ with the wedge size of 5°.

**Structure determination, refinement and analyses.** X-ray diffraction data were integrated and scaled with the HKL2000 package<sup>34</sup> and further processed with the CCP4 package<sup>35</sup>. The structure of CavAb and its derivatives were solved by molecular replacement by using an individual subunit of the NavAb structure (PDB code 3RVY) as the search template. The data sets were processed in C2 space group and there are four molecules in one asymmetric unit. We chose the I222 space group to process the data sets for initial structural determination, but we found that the bound ions were slightly off-centre with respect to the axis of the pore. Therefore, to better interpret the coordination of Ca<sup>2+</sup>, Cd<sup>2+</sup> and Mn<sup>2+</sup>, we solved the structures in the C2 space group. Crystallography and NMR System software<sup>36</sup> was used for refinement of coordinates and *B*-factors. Final models were obtained after several cycles of refinement with REFMAC<sup>37</sup> and PHENIX<sup>38</sup> and manual re-building using COOT<sup>39</sup>. The geometries of the final structural models of CavAb and its derivatives were verified using PROCHECK<sup>40</sup>. The divalent cations were identified by anomalous difference Fourier maps calculated using data collected at wavelengths of 1.75 Å for Ca<sup>2+</sup>, Cd<sup>2+</sup> and Mn<sup>2+</sup>. Detailed crystallographic data and refinement statistics for all the constructs are shown in Supplementary Table 1. All structural figures were prepared with PyMol<sup>41</sup>.

**Electrophysiology.** Wild-type NavAb expressed by infection of insect cells (High5) activates at very negative potentials (*V*<sub>1/2</sub> ~ -98 mV) and shows a strong, late use-dependent phase of slow inactivation. Mutation N49K shifts the activation curve ~75 mV to more positive potentials and abolishes the use-dependent inactivation<sup>42</sup>.

All NavAb/CavAb constructs used were made on the background of N49K mutation and showed good expression, allowing measurement of ionic currents 24–48 h after infection.

Whole-cell currents were recorded using an Axopatch 200 amplifier (Molecular Devices) with glass micropipettes (2–5 MΩ). Capacitance was subtracted and 80–90% of series resistance was compensated using internal amplifier circuitry. For reversal potential measurements, the intracellular pipette solution contained (in mM): 100 NaF, 10 NaCl, 20 HEPES-Na, 10 EGTA, pH 7.4 (adjusted with NaOH, [Na<sup>+</sup>]<sub>total</sub> = 146 mM). Extracellular solution contained (in mM) 10 CaCl<sub>2</sub>, 140 NMDG-methanesulphonate, 20 HEPES, (pH 7.4, adjusted with Ca(OH)<sub>2</sub>, [Ca<sup>2+</sup>]<sub>total</sub> = 12 mM). For Ba<sup>2+</sup> reversal potential measurements, BaCl<sub>2</sub> replaced CaCl<sub>2</sub>. Current–voltage (*I*–*V*) relationships were recorded in response to steps to voltages ranging from -100 to +70 mV in 5- or 10-mV increments from a holding potential of -100 mV. Pulses were generated and currents were recorded using Pulse software controlling an Instrutech ITC18 interface (HEKA). Data were analysed using Igor Pro 6.2 (WaveMetrics). Sample sizes were chosen to give s.e.m. values of less than 10% of peak values based on prior experimental experience.

Relative permeability values were calculated as described<sup>21</sup>. The permeability ratio was calculated as:

$$P_x/P_{Na} = \left\{ a_{Na} \left[ \exp \left( \frac{E_{rev} F}{RT} \right) \right] \left[ \exp \left( \frac{E_{rev} F}{RT} \right) + 1 \right] \right\} / 4a_x$$

In which *F*, *R*, *T* and *E*<sub>rev</sub> are Faraday constant, gas constant, absolute temperature and reversal potential, respectively. *a*<sub>x</sub> denotes the activity of the external divalent ion, *x*, (Ca<sup>2+</sup> or Ba<sup>2+</sup>) and *a*<sub>Na</sub>, the activity of intracellular sodium. The calculated activity coefficients were *γ*<sub>Ca</sub> = 0.33, *γ*<sub>Ba</sub> = 0.30, *γ*<sub>Na</sub> = 0.74. All potentials were corrected for the experimentally determined liquid junction potential.

For anomalous mole fraction and blocking experiments, the divalent (Ca<sup>2+</sup>, Cd<sup>2+</sup> and Mn<sup>2+</sup>) was diluted in 10 mM BaCl<sub>2</sub>, 140 mM NMDG-methanesulphonate and 10 mM HEPES and perfused for 2–3 min before recording a *I*–*V* curve. The peak value of the *I*–*V* curve was measured and normalized to the peak value without the divalent cation.

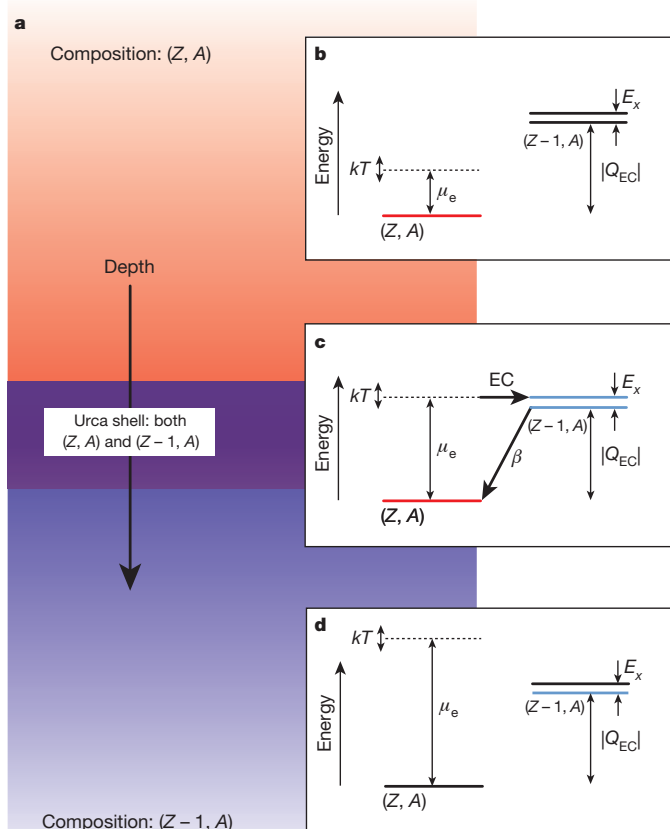
32. Faham, S. & Bowie, J. U. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *J. Mol. Biol.* **316**, 1–6 (2002).
33. Faham, S. *et al.* Crystallization of bacteriorhodopsin from bicelle formulations at room temperature. *Protein Sci.* **14**, 836–840 (2005).
34. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
35. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
36. Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
37. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
38. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
39. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
40. Laskowski, R. A., Moss, D. S. & Thornton, J. M. Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067 (1993).
41. DeLano, W. L. PyMOL molecular viewer (V.1. 2r3pre) (<http://www.pymol.org>) (2002).
42. Gamal El-Din, T. M., Martinez, G. Q., Payandeh, J., Scheuer, T. & Catterall, W. A. A gating charge interaction required for late slow inactivation of the bacterial sodium channel NavAb. *J. Gen. Physiol.* **142**, 181–190 (2013).

# Strong neutrino cooling by cycles of electron capture and $\beta^-$ decay in neutron star crusts

H. Schatz<sup>1,2,3</sup>, S. Gupta<sup>4</sup>, P. Möller<sup>2,5</sup>, M. Beard<sup>2,6</sup>, E. F. Brown<sup>1,2,3</sup>, A. T. Deibel<sup>2,3</sup>, L. R. Gasques<sup>7</sup>, W. R. Hix<sup>8,9</sup>, L. Keek<sup>1,2,3</sup>, R. Lau<sup>1,2,3</sup>, A. W. Steiner<sup>2,10</sup> & M. Wiescher<sup>2,6</sup>

The temperature in the crust of an accreting neutron star, which comprises its outermost kilometre, is set by heating from nuclear reactions at large densities<sup>1–4</sup>, neutrino cooling<sup>5,6</sup> and heat transport from the interior<sup>7–11</sup>. The heated crust has been thought to affect observable phenomena at shallower depths, such as thermonuclear bursts in the accreted envelope<sup>10,11</sup>. Here we report that cycles of electron capture and its inverse,  $\beta^-$  decay, involving neutron-rich nuclei at a typical depth of about 150 metres, cool the outer neutron star crust by emitting neutrinos while also thermally decoupling the surface layers from the deeper crust. This ‘Urca’ mechanism<sup>12</sup> has been studied in the context of white dwarfs<sup>13</sup> and type Ia supernovae<sup>14,15</sup>, but hitherto was not considered in neutron stars, because previous models<sup>1,2</sup> computed the crust reactions using a zero-temperature approximation and assumed that only a single nuclear species was present at any given depth. The thermal decoupling means that X-ray bursts and other surface phenomena are largely independent of the strength of deep crustal heating. The unexpectedly short recurrence times, of the order of years, observed for very energetic thermonuclear superbursts<sup>16</sup> are therefore not an indicator of a hot crust, but may point instead to an unknown local heating mechanism near the neutron star surface.

Continual accretion onto a neutron star pushes the ashes of surface thermonuclear burning, which is often observed as type I X-ray bursts<sup>17,18</sup>, to greater pressures and densities, at which the nuclei form a rigid lattice<sup>19</sup> known as the crust. With increasing depth, these ashes are transformed by capture of degenerate electrons into increasingly neutron-rich nuclei<sup>1–4</sup>. An electron-capture reaction— $(Z, A) + e^- \rightarrow (Z-1, A) + \nu_e$ —involves a parent nucleus  $(Z, A)$  with charge number  $Z$  and mass number  $A$  and gives rise to a daughter nucleus  $(Z-1, A)$  with the emission of an electron neutrino; this occurs at a well-defined depth, where the electron chemical potential  $\mu_e \approx |Q_{EC}| + E_x$ . Here  $Q_{EC}$  is the (negative) electron-capture  $Q$ -value (the difference between the parent and daughter ground-state masses and hence the energy needed for the reaction to occur) and  $E_x$  is the excitation energy of the lowest state in the daughter nucleus that can be populated by electron capture. In the commonly used zero-temperature approximation, the reverse  $\beta^-$ -decay reaction  $(Z-1, A) \rightarrow (Z, A) + e^- + \bar{\nu}_e$  is blocked because there is no phase space available in which to re-emit the captured electron. At finite temperature and for  $E_x < kT$ , however,  $\beta^-$  decay via the re-emission of an electron with an energy close to  $|Q_{EC}|$  is not completely blocked. As a result, the boundary between a layer containing nuclei  $(Z, A)$  and a deeper layer containing  $(Z-1, A)$  is a shell with mixed composition spanning a range of electron chemical potential  $|Q_{EC}| - kT \lesssim \mu_e \lesssim |Q_{EC}| + kT$  that corresponds to a thickness of a few metres within the neutron star crust. Inside this shell, both electron capture and its inverse,  $\beta^-$  decay, occur (see Fig. 1). If these reactions cycle back-and-forth rapidly, the



**Figure 1 | Schematic nuclear energy-level diagrams for an electron-capture/ $\beta^-$ -decay pair.** **a**, Illustration of compositional layers in the neutron star crust; **b–d**, energy level diagrams. In the shallow region above the Urca shell, where the nuclear composition has charge number  $Z$  and mass number  $A$ ,  $(Z, A)$ , the electron chemical potential  $\mu_e$  is less than  $|Q_{EC}|$ , the energy threshold for electron capture, and electron capture is energetically blocked (**b**). In the deeper region below the Urca shell,  $\mu_e > |Q_{EC}|$ : electron capture has therefore occurred, the composition consists of nuclei  $(Z-1, A)$ , and the degenerate electrons block the phase space for electron emission via  $\beta^-$  decay (**d**). In the Urca shell between these regions,  $\mu_e \approx |Q_{EC}|$ . As a result, both electron capture (EC) and  $\beta^-$  decay ( $\beta$ ) are possible (**c**), and rapid cycling between the nuclei  $(Z, A)$  and  $(Z-1, A)$  leads to a strong neutrino emissivity.

<sup>1</sup>National Superconducting Cyclotron Laboratory, Michigan State University, 640 South Shaw Lane, East Lansing, Michigan 48824, USA. <sup>2</sup>Joint Institute for Nuclear Astrophysics, 225 Nieuwland Science Hall, University of Notre Dame, Notre Dame, Indiana 46556, USA. <sup>3</sup>Department of Physics and Astronomy, Michigan State University, 567 Wilson Road, East Lansing, Michigan 48824, USA. <sup>4</sup>Indian Institute of Technology Ropar, Nangal Road, Rupnagar (Ropar), Punjab 140 001, India. <sup>5</sup>Theoretical Division, MS B214, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. <sup>6</sup>Department of Physics, 225 Nieuwland Science Hall, University of Notre Dame, Notre Dame, Indiana 46556, USA. <sup>7</sup>Departamento de Física Nuclear, Instituto de Física da Universidade de São Paulo, Caixa Postal 66318, 05315-970 São Paulo, Brazil. <sup>8</sup>Physics Division, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, Tennessee 37831-6354, USA. <sup>9</sup>Department of Physics and Astronomy, University of Tennessee, 401 Nielsen Physics Building, 1408 Circle Drive, Knoxville, Tennessee 37996-1200, USA. <sup>10</sup>Institute for Nuclear Theory, University of Washington, Physics and Astronomy Building, Box 351550, Seattle, Washington 98195-1550, USA.

**Table 1 | Electron-capture/ $\beta^-$ -decay pairs with highest cooling rates**

Electron-capture/ $\beta^-$ -decay pair		Density†	Chemical potential†	Luminosity‡
Parent	Daughter*	( $10^{10} \text{ g cm}^{-3}$ )	(MeV)	( $10^{36} \text{ erg s}^{-1}$ )
$^{29}\text{Mg}$	$^{29}\text{Na}$	4.79	13.3	24
$^{55}\text{Ti}$	$^{55}\text{Sc}$ , $^{55}\text{Ca}$	3.73	12.1	11
$^{31}\text{Al}$	$^{31}\text{Mg}$	3.39	11.8	8.8
$^{33}\text{Al}$	$^{33}\text{Mg}$	5.19	13.4	8.3
$^{56}\text{Ti}$	$^{56}\text{Sc}$	5.57	13.8	3.5
$^{57}\text{Cr}$	$^{57}\text{V}$	1.22	8.3	1.6
$^{57}\text{V}$	$^{57}\text{Ti}$ , $^{57}\text{Sc}$	2.56	10.7	1.6
$^{63}\text{Cr}$	$^{63}\text{V}$	6.82	14.7	0.97
$^{105}\text{Zr}$	$^{105}\text{Y}$	3.12	11.2	0.92
$^{59}\text{Mn}$	$^{59}\text{Cr}$	0.945	7.6	0.88
$^{103}\text{Sr}$	$^{103}\text{Rb}$	5.30	13.3	0.65
$^{96}\text{Kr}$	$^{96}\text{Br}$	6.40	14.3	0.65
$^{65}\text{Fe}$	$^{65}\text{Mn}$	2.34	10.3	0.60
$^{65}\text{Mn}$	$^{65}\text{Cr}$	3.55	11.7	0.46

\* The listing of two electron-capture daughter isotopes means that two subsequent reaction pairs occur in the same layer.

† The transition always occurs at the specified electron chemical potential. The density, which is for a composition consisting of nuclei with a single mass number  $A$ , will only be approximate for an arbitrary composition.

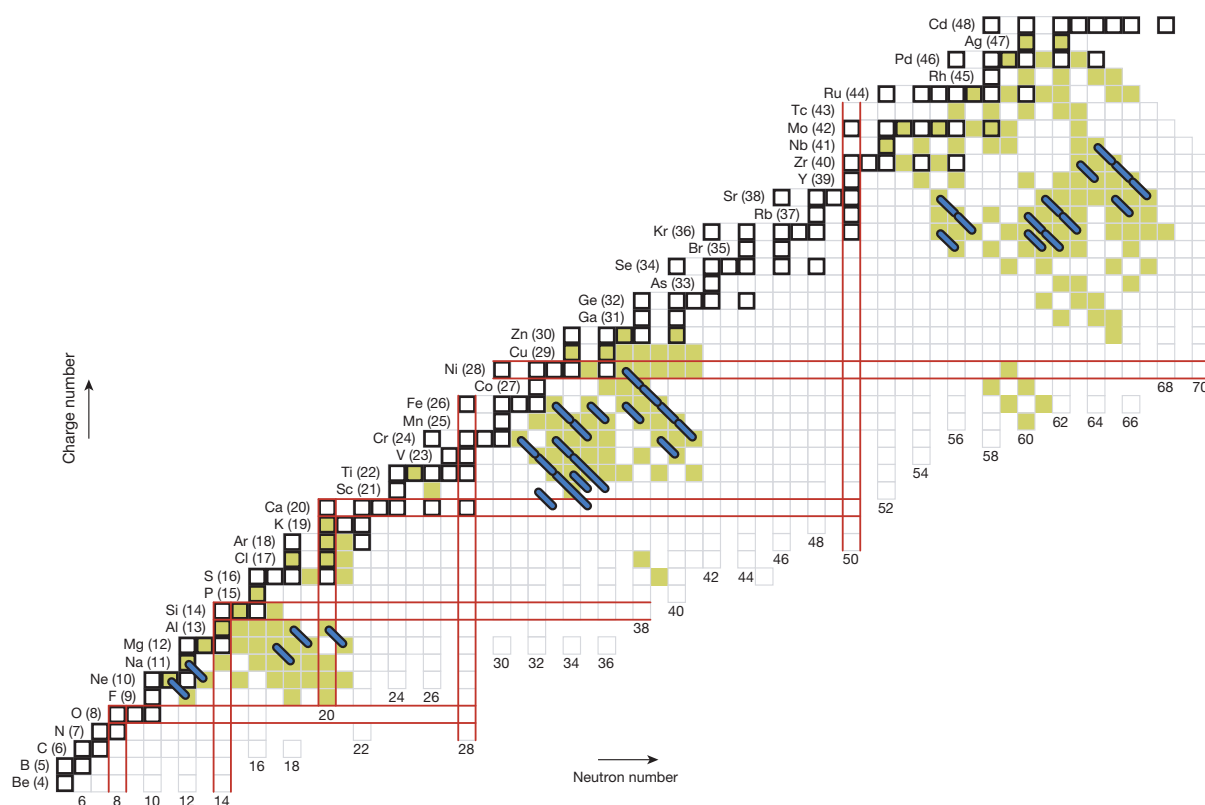
‡ The cooling luminosity  $L_\nu$  scales with temperature  $T$ , local gravitational acceleration  $g$ , neutron star radius (in the local rest frame)  $R$ , and mass fraction  $X$  of the respective electron-capture/ $\beta^-$ -decay pair as  $L_\nu \propto X R^2 g^{-1} T^6$ . The temperature scaling assumes  $E_x \ll kT$ . For further details, see Supplementary Information section 1. The values for  $L_\nu$  we quote here are for  $T = 0.51 \text{ GK}$ ,  $g = 1.85 \times 10^{14} \text{ cm s}^{-2}$ ,  $R = 12 \text{ km}$  and  $X = 1$ . The existence of the  $^{56}\text{Ti}$ - $^{56}\text{Sc}$  electron-capture/ $\beta^-$ -decay pair depends strongly on nuclear masses. In all other cases nuclear-physics-related uncertainties of the predicted luminosities are of the order of a factor of 3–4 (see Supplementary Information section 5).

result is a strong neutrino emission, known as an Urca process<sup>12</sup>, that cools the neutron star crust.

Such Urca shells are thought to also operate in white dwarfs<sup>13</sup>, type Ia supernovae<sup>14,15</sup> and stellar ONeMg cores producing electron-capture

supernovae<sup>20</sup>. But the effect has not been considered in the context of accreting neutron stars. Most earlier models of accreted crusts<sup>1,2,4</sup> were computed in the zero-temperature limit, in which the composition switches sharply at the energetic thresholds with no available phase space for cycling; indeed, in this limit, only one nuclide of a reaction pair is present at a given depth. Urca shell cooling relies on phase space unblocking at finite temperature and the presence of both reaction pair nuclides in the shell. More recent reaction network calculations<sup>3</sup> did not include  $\beta^-$ -decays as they were not considered to be important, and any Urca cycling was estimated to be negligible. The importance of Urca shell cooling is revealed here through the use of a full reaction network that includes both electron capture and  $\beta^-$  decay on an equal footing, that takes into account the rates of subsequent reactions that deplete the electron-capture/ $\beta^-$ -decay pairs, and that follows the evolution in time of a fluid element as it is pushed through a reaction shell.

In order for this cycling of electron capture and  $\beta^-$  decays between two nuclear species to occur, the nuclei involved must satisfy two conditions. First, the transitions must proceed between low-lying states ( $E_x \lesssim kT$  is required for both the electron capture and the  $\beta^-$  decay). In addition, within an electron-capture/ $\beta^-$ -decay pair, the nucleus undergoing  $\beta^-$  decay must not have a strong electron-capture branch, as these electron captures would remove nuclei from the Urca cycle, thereby reducing its effect or eliminating it entirely. The cooling rate depends on the strength of the transition (the  $ft$  value, which is proportional to the matrix element connecting the parent and daughter states) and the energy threshold; the integration over the available phase space produces a characteristic  $T^5$  scaling with temperature<sup>13</sup>. The formation of Urca shells with large cooling rates is enabled by strong nuclear deformations that tend to spread nuclear electron-capture strength to lower



**Figure 2 | Electron-capture/ $\beta^-$ -decay pairs on a chart of the nuclides.** The thick blue lines denote electron-capture/ $\beta^-$ -decay pairs that would generate a strong neutrino luminosity in excess of  $5 \times 10^{34} \text{ erg s}^{-1}$  at  $T = 0.51 \text{ GK}$  for a composition consisting entirely of the respective electron-capture/ $\beta^-$ -decay pair. They largely coincide with regions where allowed electron-capture and  $\beta^-$ -decay transitions are predicted to populate low-lying states and subsequent electron capture is blocked (shaded squares, see also the discussion

in ref. 3). These are mostly regions between the closed neutron and proton shells (pairs of horizontal and vertical red lines), where nuclei are significantly deformed (see Supplementary Information section 4). Nuclides that are  $\beta^-$ -stable under terrestrial conditions are shown as squares bordered by thicker lines. Nuclear charge numbers are indicated in parentheses next to element symbols.

excited states, thereby lowering  $E_x$  (ref. 21; see Extended Data Fig. 1). There are a number of electron-capture/ $\beta^-$ -decay pairs that fulfil these conditions for forming fast-cooling Urca shells (see Table 1 and Fig. 2). The degree to which these shells are activated in a neutron star crust depends on the initial composition produced by thermonuclear burning on the neutron star surface. Because electron capture in the outer crust preserves the mass number  $A$ , the abundance of an electron-capture/ $\beta^-$ -decay pair, and therefore its absolute neutrino luminosity, is set by the abundance of nuclei with the same mass number in the ashes of the surface thermonuclear burning. As is evident from Fig. 3, neutrino cooling by Urca shells is by far the dominant neutrino emission process in the crust for typical crust compositions.

The greatly enhanced crust neutrino emissivity at rather shallow depths changes the long-standing assumption that rapidly accreting neutron stars have a significant luminosity from deep crustal heating, which directly influences thermonuclear burning in their accreted envelopes. In the absence of crust Urca shells, if the crust has a low thermal conductivity, and if the core neutrino emissivity is weak, then deep crustal heating would generate significant heat flow towards the neutron star surface<sup>22</sup>. Models of thermonuclear bursts<sup>11,17</sup> use this emergent luminosity from the deep crust as a boundary condition, which sets in part the ignition density and temperature. The presence of strongly temperature-sensitive Urca cycles limits the temperature at the location of the shells, however, and may even require an inward directed luminosity from the accreted envelope. Even for conditions in the deep

crust that are favourable for sending a large heat flux to the surface, the Urca shells re-emit this heat as neutrinos, thereby preventing it from reaching the surface layers (see Supplementary Information section 3).

To establish the robustness of our conclusions with respect to nuclear physics uncertainties, we used two different mass models, namely FRDM<sup>23</sup> and HFB-21<sup>24</sup>. The use of FRDM masses instead of those from HFB-21 reduces the Urca-shell neutrino luminosity by 90% for a superburst ash composition (see Supplementary Information section 5 for details). For both mass models, the temperature at the superburst ignition depth is  $<5 \times 10^8$  K if the Urca shell neutrino emissivity is included; and for both mass models the temperature has a significant local minimum at the location of the Urca shell (see Extended Data Fig. 2).

This has important implications for the ignition of superbursts, which are thought to be triggered by the unstable thermonuclear reaction  $^{12}\text{C} + ^{12}\text{C}$ . The observed recurrence times, of the order of one year, are much shorter than predictions of current models<sup>7,8,10</sup>, which indicates that the temperature at the ignition depth is underestimated. The presence of Urca shells implies that this observation does not point to an unexpectedly hot crust<sup>3,10</sup>. Because of the Urca shells, we conclude that the standard carbon-ignition scenario for superbursts requires a powerful, as yet unknown, heat source that operates at surprisingly shallow densities  $\lesssim 10^{10} \text{ g cm}^{-3}$  very near the carbon ignition layer. Alternatively other, more exotic, mechanisms would have to be found to ignite and power superbursts.

Urca shell cooling therefore forces fundamental changes in current superburst models. Realistic ignition conditions must now include a strong localized heat source at a depth close to that of ignition, with a strong heat flux flowing inwards into the Urca shell cooling layer. The resulting temperature profile at ignition is therefore dramatically different from that assumed previously, which may alter predicted light curves. In addition, during the explosion, the temperature at the ignition depth rises to  $\sim 10^9$  K. Heat from this layer will diffuse inward; on the basis of the thermal diffusion timescale in the neutron star crust<sup>25</sup>, we estimate that in the absence of any neutrino emission, the temperature would rise to  $\sim 10^9$  K at the depth of the Urca shell within a day following ignition. The presence of a strong 'heat sink' at that depth, however, will prevent the deeper layers from rising in temperature and will, therefore, force the observed light curve to decay faster than expected over timescales of roughly one day. Current superburst observations on this timescale are rare and provide data of limited quality<sup>26</sup>, but a dedicated programme of superburst follow-up observations with current instrumentation could address this problem. Detailed simulations, which are beyond the scope of this Letter, are required to quantify the effect of Urca shell cooling on superburst light curves.

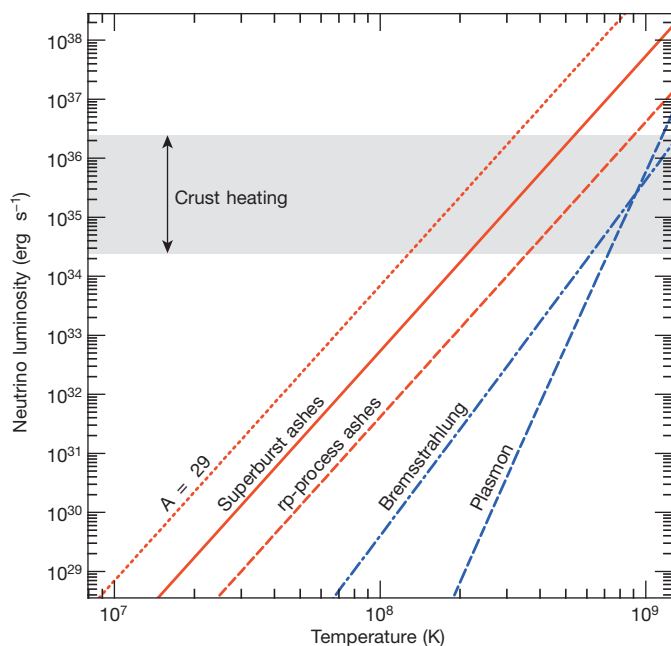
Another observational signature might be found in neutron star cooling following an accretion outburst. Unlike crustal heating, the rate of Urca shell cooling does not scale with accretion rate, but rather depends only on temperature. Cooling will therefore continue in transiently accreting neutron stars once accretion has ceased, and might affect observations of the cooling crusts in the hottest of these systems, such as XTE J1701–462<sup>27</sup>; in such systems, the Urca shell neutrino cooling rate is comparable to typical photon luminosities of  $10^{32}$ – $10^{33} \text{ erg s}^{-1}$  for typical initial crust temperatures of  $(1\text{--}3) \times 10^8$  K.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 March; accepted 1 October 2013.

Published online 1 December 2013.

1. Sato, K. Nuclear compositions in the inner crust of neutron stars. *Prog. Theor. Phys.* **62**, 957–968 (1979).
2. Haensel, P. & Zdunik, J. L. Non-equilibrium processes in the crust of an accreting neutron star. *Astron. Astrophys.* **227**, 431–436 (1990).
3. Gupta, S., Brown, E. F., Schatz, H., Möller, P. & Kratz, K.-L. Heating in the accreted neutron star ocean: implications for superburst ignition. *Astrophys. J.* **662**, 1188–1197 (2007).



**Figure 3 | Neutrino luminosities in the accreted neutron star crust.** For purposes of comparison, estimates of the neutrino luminosities were obtained by integrating the corresponding emissivities over a neutron star crust with a constant local temperature, following the methodology of ref. 10. Although an actual neutron star crust is not isothermal, the temperature variation across the crust is not large (typically less than a factor of 2). The Urca-shell luminosities were calculated for superburst ashes<sup>26</sup> (solid red line), X-ray burst ashes produced by the rapid proton-capture (rp) process<sup>17</sup> (dashed red line), and a pure  $A = 29$  composition (dotted red line) to demonstrate the maximum effect. Also shown are the neutrino luminosities from electron–nucleus bremsstrahlung (blue dot-dashed line) and plasmon decay (blue dashed line). For comparison, we show estimates for the total crustal heating (shaded band). This is based on a local heating rate of  $(1.9 \text{ MeV/nucleon}) \times \dot{M}$  for a range of accretion rates  $10^{16} \text{ g s}^{-1} \lesssim \dot{M} \lesssim 10^{18} \text{ g s}^{-1}$ , which is representative for observed neutron stars. Urca shells dominate the neutrino luminosity from the crust and can balance the crust heating at moderate temperatures  $\gtrsim 2 \times 10^8$  K. The temperature scaling assumes  $E_x \ll kT$  (see Supplementary Information section 1).



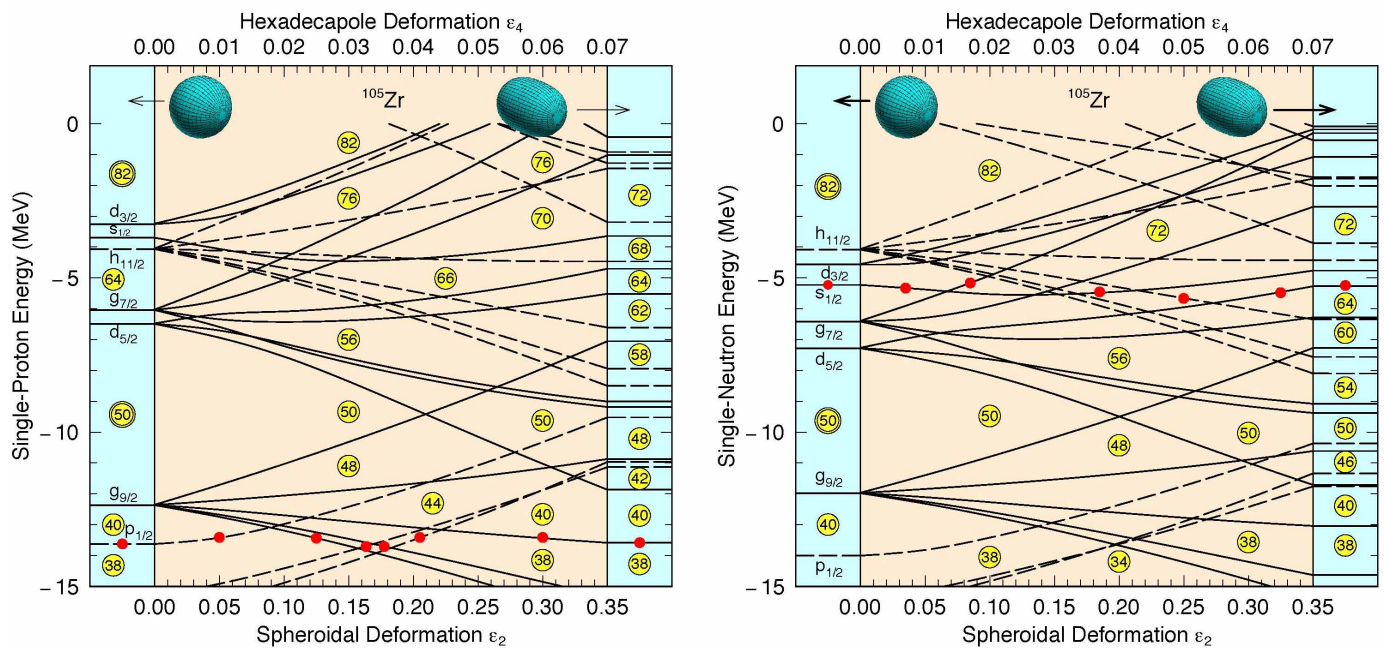
4. Haensel, P. & Zdunik, J. L. Models of crustal heating in accreting neutron stars. *Astron. Astrophys.* **480**, 459–464 (2008).
5. Yakovlev, D. G., Kaminker, A. D., Gnedin, O. Y. & Haensel, P. Neutrino emission from neutron stars. *Phys. Rep.* **354**, 1–155 (2001).
6. Steiner, A. W. & Reddy, S. Superfluid response and the neutrino emissivity of neutron matter. *Phys. Rev. C* **79**, 015802 (2009).
7. Cumming, A. & Bildsten, L. Carbon flashes in the heavy-element ocean on accreting neutron stars. *Astrophys. J.* **559**, L127–L130 (2001).
8. Strohmayer, T. E. & Brown, E. F. A remarkable 3 hour thermonuclear burst from 4U 1820–30. *Astrophys. J.* **566**, 1045–1059 (2002).
9. Brown, E. F. Superburst ignition and implications for neutron star interiors. *Astrophys. J.* **614**, L57–L60 (2004).
10. Cumming, A., Macbeth, J., in 't Zand, J. J. M. & Page, D. Long type I X-ray bursts and neutron star interior physics. *Astrophys. J.* **646**, 429–451 (2006).
11. Keek, L. & Heger, A. Multi-zone models of superbursts from accreting neutron stars. *Astrophys. J.* **743**, 189–203 (2011).
12. Gamow, G. & Schoenberg, M. Neutrino theory of stellar collapse. *Phys. Rev.* **59**, 539–547 (1941).
13. Tsuruta, S. & Cameron, A. G. W. Urca shells in dense stellar interiors. *Astrophys. J.* **7** (Suppl.), 374–406 (1970).
14. Paczyński, B. Carbon ignition in degenerate stellar cores. *Astrophys. J.* **11**, 53–55 (1972).
15. Woosley, S. E. & Weaver, T. A. The physics of supernova explosions. *Annu. Rev. Astron. Astrophys.* **24**, 205–253 (1986).
16. Keek, L. & in 't Zand, J. J. M. On burning regimes and long duration X-ray bursts. In *Proc. 7th INTEGRAL Workshop* (eds Brandt, S., Westergaard, N. J. & Lund, N.) PoS(Integral08)032 (Proceedings of Science, 2008).
17. Woosley, S. *et al.* Models for type I X-ray bursts with improved nuclear physics. *Astrophys. J.* **151** (Suppl.), 75–102 (2004).
18. Galloway, D. K., Muno, M. P., Hartman, J. M., Psaltis, D. & Chakrabarty, D. Thermonuclear (type-I) X-ray bursts observed by the Rossi X-ray Timing Explorer. *Astrophys. J.* **179** (Suppl.), 360–422 (2008).
19. Horowitz, C. J. & Berry, D. K. Structure of accreted neutron star crust. *Phys. Rev. C* **79**, 065803 (2009).
20. Jones, S. *et al.* Advanced burning stages and fate of 8–10  $M_{\odot}$  stars. *Astrophys. J.* **772**, 150–163 (2013).
21. Kruumlind, J. & Möller, P. Calculation of Gamow-Teller  $\beta$ -strength functions in the rubidium region in the RPA approximation with Nilsson-model wave functions. *Nucl. Phys. A* **417**, 419–446 (1984).
22. Brown, E. F. Nuclear heating and melted layers in the inner crust of an accreting neutron star. *Astrophys. J.* **531**, 988–1002 (2000).
23. Möller, P., Nix, J. R., Myers, W. D. & Swiatecki, W. J. Nuclear ground-state masses and deformations. *At. Data Nucl. Data Tables* **59**, 185–381 (1995).
24. Goriely, S., Chamel, N. & Pearson, J. M. Further explorations of Skyrme-Hartree-Fock-Bogoliubov mass formulas. XII. Stiffness and stability of neutron-star matter. *Phys. Rev. C* **82**, 035804 (2010).
25. Brown, E. F. & Cumming, A. Mapping crustal heating with the cooling light curves of quasi-persistent transients. *Astrophys. J.* **698**, 1020–1032 (2009).
26. Keek, L., Heger, A. & in 't Zand, J. J. M. Superburst models for neutron stars with hydrogen-and helium-rich atmospheres. *Astrophys. J.* **752**, 150–162 (2012).
27. Page, D. & Reddy, S. Forecasting neutron star temperatures: predictability and variability. Preprint at <http://arXiv.org/abs/1307.4455> (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This project was funded by NSF grants PHY 08-22648 (Joint Institute for Nuclear Astrophysics) and PHY 06-06007. A.W.S. was supported by INT DOE grant DE-FG02-00ER41132. E.F.B. was supported by NSF grant AST 11-09176. P.M. was supported by the National Nuclear Security Administration of the US Department of Energy at Los Alamos National Laboratory under contract no. DE-AC52-06NA25396. We thank D.M. Yakovlev, P. Shternin and S. Reddy for discussions and comments on the manuscript.

**Author Contributions** H.S. calculated crust models, analysed data and prepared the manuscript. S.G. developed and implemented the phase space calculation. S.G. and P.M. calculated the weak transition rates. E.F.B. calculated crust temperature profiles and assisted with writing the manuscript. A.T.D. computed the temperature scaling of the neutrino emission. L.K. calculated superburst models. M.B., W.R.H. and R.L. wrote model code. All authors contributed to the interpretation of the results, and contributed to or commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.S. ([schatz@nsl.msu.edu](mailto:schatz@nsl.msu.edu)).

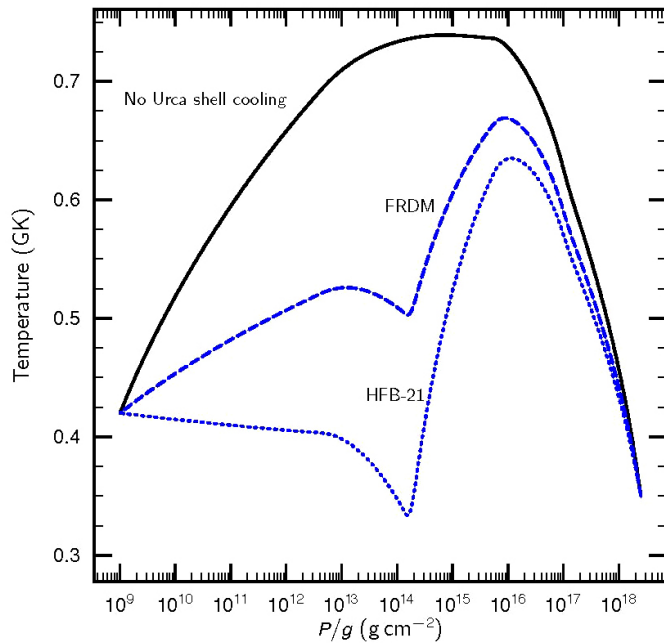


**Extended Data Figure 1 | Calculated proton and neutron single-particle energy levels in  $^{105}\text{Zr}$  as functions of nuclear deformation.** Left panel, proton levels; right panel, neutron levels. The 40 protons and 65 neutrons in  $^{105}\text{Zr}$  fill all levels up to the Fermi levels corresponding to these nucleon numbers in the two diagrams (red dots). Levels corresponding to even parity are shown as solid lines, those corresponding to odd parity as dashed lines. Shell gaps are characterized by a particularly large separation in energy between two adjacent single-particle levels. The numbers of protons or neutrons that occupy levels up to the shell gap are indicated by circled numbers. The single-particle levels are shown for a spherical nucleus in spectroscopic standard notation (left side of each panel), and for a deformation near the calculated ground-state

shape of  $^{105}\text{Zr}$  with quadrupole and hexadecapole shape-parameter values  $\varepsilon_2 = 0.333$  and  $\varepsilon_4 = 0.06$ , respectively<sup>23</sup> (right side of each panel). The middle section of each panel shows the change in level energies as  $\varepsilon_2$  and  $\varepsilon_4$  change from spherical values  $\varepsilon_2 = \varepsilon_4 = 0$  to deformed values<sup>28</sup>. The well-known “magic numbers” 50 and 82 corresponding to particularly large gaps stand out at zero deformation<sup>29</sup>. When the nuclear shape becomes deformed, the spherical shell gaps disappear resulting in a large density of levels in the vicinity of the Fermi level. This gives rise to a large number of states at low excitation in  $^{105}\text{Zr}$ . Some of these states can be populated by strong  $\beta^-$  decay transitions from the ground state of  $^{105}\text{Y}$ . The situation is similar for the electron capture on  $^{105}\text{Zr}$  into deformed  $^{105}\text{Y}$ .

28. Nilsson, S. G. Binding state of individual nucleons in strongly deformed nuclei. *Mat.-fys. Meddr.* **29**, 1–69 (1955).

29. Mayer, M. G. Nuclear configurations in the spin-orbit coupling model. II. Theoretical considerations. *Phys. Rev.* **78**, 22–23 (1950).



**Extended Data Figure 2 | Temperature as a function of depth in the accreted neutron star crust for different Urca shell cooling strengths.** Here we use  $P/g \approx \int \rho \, dz$  as a proxy for depth, where  $P$  is the pressure,  $g$  the local gravitational acceleration,  $\rho$  the mass density and  $z$  the spatial depth coordinate. As a baseline model, we fix the temperature to be  $T = 0.42$  GK at  $P/g = 10^9 \text{ g cm}^{-2}$  and  $T = 0.35$  GK at the crust–core transition. In the absence of Urca shell cooling, the peak local temperature reaches 0.73 GK (solid curve) with the temperature at the superburst ignition depth ( $P/g \approx 10^{12} \text{ g cm}^{-2}$ ) being 0.66 GK. With the addition of cooling using the HFB-21 mass model and a superburst ash composition (blue dotted line), a local temperature minimum,  $T = 0.33$  GK, appears at the location of the Urca shell. Indeed, for these conditions the temperature at the Urca shell is lower than that at the upper boundary, so that a temperature inversion develops. Even for the much lower Urca shell emissivity of the FRDM mass model (blue dashed line), the temperature at the depth of the superburst ignition is  $\lesssim 5 \times 10^8$  K, which is inconsistent with typical superburst ignition conditions<sup>10</sup>. For both mass models, the temperature has a local minimum at the location of the Urca shell. The steady-state cooling luminosity from the shell is  $2 \times 10^{35} \text{ erg s}^{-1}$  for the HFB-21 mass model and  $1.4 \times 10^{35} \text{ erg s}^{-1}$  for the FRDM mass model. As a result, the Urca shell thermally decouples the envelope of light elements from the heating in the deeper crust.

# A featureless transmission spectrum for the Neptune-mass exoplanet GJ 436b

Heather A. Knutson<sup>1</sup>, Björn Benneke<sup>1,2</sup>, Drake Deming<sup>3</sup> & Derek Homeier<sup>4</sup>

**GJ 436b is a warm—approximately 800 kelvin—exoplanet that periodically eclipses its low-mass (half the mass of the Sun) host star, and is one of the few Neptune-mass planets that is amenable to detailed characterization. Previous observations<sup>1–3</sup> have indicated that its atmosphere has a ratio of methane to carbon monoxide that is  $10^5$  times smaller than predicted by models for hydrogen-dominated atmospheres at these temperatures<sup>4,5</sup>. A recent study proposed that this unusual chemistry could be explained if the planet's atmosphere is significantly enhanced in elements heavier than hydrogen and helium<sup>6</sup>. Here we report observations of GJ 436b's atmosphere obtained during transit. The data indicate that the planet's transmission spectrum is featureless, ruling out cloud-free, hydrogen-dominated atmosphere models with an extremely high significance of  $48\sigma$ . The measured spectrum is consistent with either a layer of high cloud located at a pressure level of approximately one millibar or with a relatively hydrogen-poor (three per cent hydrogen and helium mass fraction) atmospheric composition<sup>7–9</sup>.**

We observed four transits of the Neptune-mass planet GJ 436b on 26 October 2012 Universal Time (UT), 29 November 2012, 10 December 2012 and 2 January 2013 using the red grism (1.2–1.6  $\mu\text{m}$ ) on the Hubble Space Telescope (HST) Wide Field Camera 3 instrument. These data were obtained in a new scanning mode<sup>10,11</sup> with a scan rate of 0.99'' per second, which allowed us to achieve approximately a factor-of-twenty improvement in the orbit-averaged efficiency compared to staring-mode observations<sup>12</sup>. Each visit spanned four HST orbits with an integration time of 7.6 s per exposure. We extracted the spectra from the raw images using the template-fitting technique described in a previous study<sup>11</sup>, and provide additional details of our reduction in the Methods.

We fitted the four wavelength-integrated (white-light) transit curves simultaneously<sup>13</sup> while accounting for detector effects (see discussion in the Methods), to determine values for the planet's orbital inclination  $i$ , the planet–star radius ratio  $R_p/R_*$ , the ratio of the planet's semi-major axis  $a$  to the stellar radius  $R_*$ , and the centre-of-transit time  $T_c$ . We set the uncertainties on each measurement equal to the standard deviation of the residuals from our best-fit solution for that visit and evaluated the uncertainties on our best-fit parameters using the covariance matrix from our Levenberg–Marquardt least-squares minimization, a Markov Chain Monte Carlo analysis, and a residual permutation technique that better accounts for the presence of time-correlated noise in the data<sup>14,15</sup>. The residual permutation approach results in uncertainties that are a factor of 1.5–2 larger than both the covariance matrix and the Markov Chain Monte Carlo errors for all of our fitting parameters, and we took those to be our final errors.

Our best-fit parameters for the white-light transits are given in Table 1, and the normalized transit light curves are shown in Fig. 1. We see no evidence for transit depth variations comparable to those reported with the Spitzer data, and we do not detect any star spot occultations in our transit light curves that are similar to those observed for HD 189733b (refs 16 and 17). We next determined the differential wavelength-dependent transit depths in twenty-eight bins spanning

wavelengths between 1.14  $\mu\text{m}$  and 1.65  $\mu\text{m}$ , as described in Fig. 2 and the Methods. The resulting transmission spectrum is shown in Fig. 2, with error bars that include both the uncertainties in the measured transit depth and in the stellar limb-darkening models.

We interpreted the transmission spectrum using a variation of the Bayesian atmospheric retrieval framework described in a previous study<sup>18</sup>; we provide a summary of this approach in the Methods. We find that we obtain the best match to our data using models with either high-altitude clouds or a relatively hydrogen-poor atmosphere with a reduced scale height and correspondingly small absorption features. At low metallicities our model requires a haze or cloud layer at pressures below 10 mbar, because the large scale height of these models otherwise leads to strong spectral signatures from molecular absorption (for example,  $\text{H}_2\text{O}$ ,  $\text{CH}_4$ ,  $\text{CO}$  or  $\text{CO}_2$ ). At higher metallicities the scale height of the atmosphere is reduced, and no clouds are needed to produce an effectively flat transmission spectrum. Our conclusions are similar to those obtained for the transiting super-Earth GJ 1214b (ref. 12), although in this case new upper limits on the planet's transmission spectrum indicate that it must have a high cloud layer even if the atmosphere is metal-rich<sup>19</sup>. GJ 436b is four times more massive with a nearly identical average density and therefore seems a less obvious candidate for a hydrogen-poor atmosphere. The 68% ( $1\sigma$ ) Bayesian credible region (confidence region) extends along a curve from hydrogen-dominated models with a high-altitude cloud layer between 0.01 mbar and 4 mbar to high-metallicity models that may or may not contain clouds.

For atmospheres reflecting the Sun's bulk composition (primarily hydrogen and helium with small amounts of heavier elements), a cloud or haze layer at 1 mbar that is optically thick for slant viewing geometries represents the best fit to the data. Zinc sulphide and potassium chloride are both plausible candidates for the composition of the cloud, because the condensation curves of these substances can readily cross the pressure–temperature profile at the millibar level in GJ 436b's atmosphere<sup>6,20</sup>. A recent study<sup>20</sup> of the super-Earth GJ 1214b, which is also a

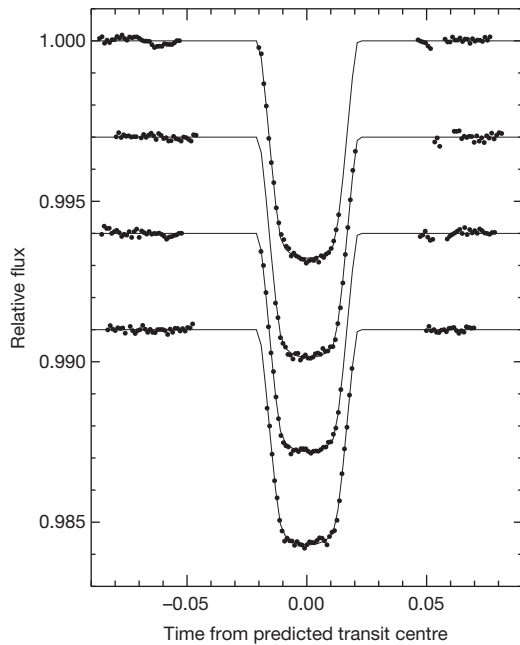
**Table 1 | Best-fit  $1\sigma$  transit parameters from white-light curves**

Visit date	$R_p/R_*$	$T_c$ (BJD <sub>TDB</sub> )
26 October 2012 UT	0.08349(33)	2456226.69131(11)
29 November 2012 UT	0.08413(26)	2456261.06211(10)
10 December 2012 UT	0.08372(31)	2456271.63758(10)
2 January 2013 UT	0.08310(27)	2456295.43270(7)
Average inclination, $i$ (°)	86.774(30)	
Average $a/R_*$	14.41(10)	

We find that our estimates (averaged over all dates) for the orbital inclination  $i$  and the ratio of the semi-major axis  $a$  to the stellar radius  $R_*$  are consistent at the  $3\sigma$  level with our previously published 8- $\mu\text{m}$  Spitzer transit observations<sup>3</sup>, and that our individual values for the planet–star radius ratio  $R_p/R_*$  are mutually consistent at the  $3\sigma$  level. We derive a new transit ephemeris with a centre-of-transit time  $T_0 = 2456295.431924(45)$  BJD<sub>TDB</sub> (where  $T_0$  is the best-fit zero point for the linear ephemeris fit) and an orbital period of  $P = 2.64389782(8)$  days by combining our data with previous studies<sup>3</sup>; our new data extend the current baseline for this object by almost 4 years. Errors are given in parentheses after the best-fit value for the orbital period, with the number  $N$  of significant digits in the quoted errors corresponding to the last  $N$  significant digits of the best-fit value. For example, the stated orbital period corresponds to a value of  $2.64389785 \pm 0.00000008$  days using standard notation. BJD, barycentric Julian date for the measured centre-of-transit times. To convert from TDB to Coordinated Universal Time (UTC) time standards, simply subtract 67.18 s from the reported centre-of-transit times,  $T_c$ .

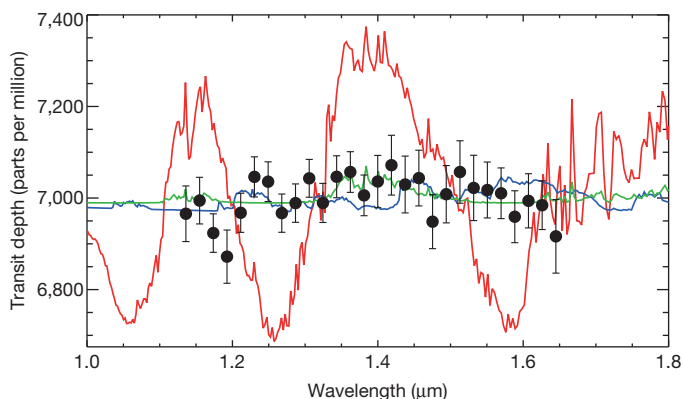
<sup>1</sup>Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. <sup>2</sup>Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>3</sup>Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. <sup>4</sup>Centre de Recherche Astrophysique de Lyon, 69364 Lyon, France.



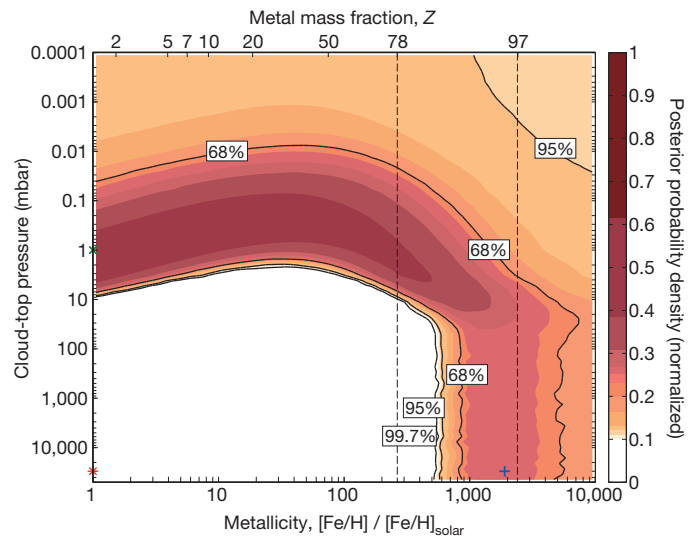


**Figure 1 | White-light transit curves for the four individual visits.** Data are vertically offset for clarity. Transits were observed on the following dates (from top to bottom): 26 October 2012 UT, 29 November 2012 UT, 10 December 2012 UT and 2 January 2013 UT. Normalized data with the first orbit trimmed and instrumental effects removed are shown as black filled circles. Best-fit model light curves are shown as black lines. The data consist of three spacecraft orbits with durations of approximately 1.5 hours each; there is a gap during each orbit where the spacecraft passes behind the Earth and the target is no longer visible.

good candidate for having these cloud species, indicates that a solar-composition atmosphere would not have sufficient amounts of condensable material to form optically thick clouds. If the atmospheric metallicity (defined as the abundances of elements heavier than hydrogen and helium, H and He) is enhanced above this level, then such clouds could easily explain this planet's flat transmission spectrum<sup>20</sup>. Alternatively, photochemical haze production could lead to an opaque cloud layer at the millibar level, although these models probably also require an enhanced atmospheric metallicity.



**Figure 2 | Averaged transmission spectrum for GJ 436b.** Black filled circles indicate the error-weighted mean transit depth in each bandpass, with the plotted uncertainties calculated as the sum in quadrature of the  $1\sigma$  standard deviation measurement errors and the systematic uncertainties from stellar limb-darkening models. We show three models for comparison, including a solar-metallicity cloud-free model (red line), a hydrogen-poor 1,900-times-solar model (blue line) and a solar-metallicity model with optically thick clouds at 1 mbar (green line).



**Figure 3 | Joint constraints on cloud-top pressure versus atmospheric metallicity.** The coloured shading indicates the normalized probability density as a function of metallicity (defined as the relative abundance of elements heavier than H and He) and cloud top pressure derived using our Bayesian atmosphere retrieval framework<sup>18,30</sup>. The black contours show the 68%, 95% and 99.7% Bayesian credible regions. We rule out a hydrogen-dominated, cloud-free, solar-metallicity atmosphere with a significance of  $48\sigma$ . Colour-matching markers indicate the three models plotted in Fig. 2, and vertical dashed lines indicate constraints on the planet's composition from measurements of its average density.

Previous studies have placed constraints on the possible bulk compositions for GJ 436b using the planet's measured mass and radius, the estimated age of the system, and models of planet formation and migration<sup>7–9,21–23</sup>. A recent survey of the published literature for this planet indicated that current models are consistent with bulk metallicities between 230 and 2,000 times solar, depending on the assumed ratio of rock to ice, the distribution of metals between the core and the envelope, the interior temperature of the planet, and other related factors<sup>6</sup>. This corresponds to a H/He mass fraction of approximately 3%–22%. On the basis of this analysis we conclude that an atmospheric metallicity of 1,900 times solar is consistent with current constraints for the planet's bulk composition, although it is very close to the upper end of this range. Mass loss appears to be minimal for GJ 436b under present conditions<sup>24–26</sup>, but it is possible that the higher ultraviolet and X-ray fluxes expected for young stars could have resulted in the loss of some atmospheric hydrogen very early in the planet's history<sup>27,28</sup>. Although a recent study<sup>29</sup> argued that such mass loss is unlikely to result in significant depletion of hydrogen relative to other elements, additional modelling work is still needed to provide a more definitive resolution to this question.

There are several potential ways to distinguish between cloudy and high-atmospheric-metallicity scenarios for GJ 436b's atmosphere. An unambiguous solution would be to obtain more precise, moderate-resolution transmission spectroscopy<sup>30</sup> capable of detecting near-infrared absorption features and directly constraining the mean molecular weight. We emphasize, however, that the apparent variations in the planet's measured transit depth from one epoch to the next (see Methods) make simultaneous measurements essential for robust constraints on this planet's transmission spectrum. A hydrogen-dominated atmosphere with a high cloud or haze layer should exhibit attenuated water absorption features, which could be distinguished from the intrinsically weaker features of high-metallicity atmospheres on the basis of the steepness of the wings of the absorption lines. Similarly, a detection of the Rayleigh scattering slope in the planet's visible-light transmission spectrum could directly constrain both the mean molecular weight and the amount of spectrally inactive gas (that is, H and He or haze particles) present in

the atmosphere<sup>20,30</sup>. Alternatively, one could differentiate between hydrogen-dominated and high-metallicity atmospheres by measuring the relative abundances of CO, CO<sub>2</sub>, methane and water. These relative abundances could then be compared to different chemical models for GJ 436b's atmosphere that might rule out the presence of significant molecular hydrogen. Improved constraints on the atmospheric chemistry from secondary eclipse spectroscopy<sup>1,2</sup>, which is less sensitive to high-altitude clouds and stellar activity, could also help to restrict the range of plausible atmospheric compositions in the limit of a well-mixed atmosphere (that is, no significant compositional gradients between the day and night sides). Lastly, improved estimates for the stellar mass and radius would help to reduce the uncertainties in the corresponding planetary values and hence better constrain its mean density.

## METHODS SUMMARY

We calculate this transmission spectrum as follows: first we determine the difference between our extracted spectrum and a best-fit template spectrum at each pixel position and create a time series of the residuals. We then fit this time series with a model consisting of the difference between the white-light transit curve and a transit light curve with a freely varying planet–star radius ratio. We also include a linear function of time to account for the first order of any remaining instrumental trends. We compare the errors on the planet–star radius ratio from the Levenberg–Marquardt covariance matrix and the residual permutation method and take the larger of the two as our final uncertainty; they typically agree to within 10%. We then average the planet–star radius ratios in four-pixel-wide segments to create our final transmission spectrum for each visit, in which we select our wavelength range to exclude the low-illumination regions at the edges of the spectrum. We calculate uncertainties on each bin as the average of the errors for the four individual radius ratios to account for the four-pixel-wide Gaussian smoothing function we applied to the raw spectra before fitting the template spectrum. We combine the data from our four visits by taking the error-weighted mean of the transit depths in each wavelength bin.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 24 July; accepted 18 November 2013.**

- Stevenson, K. B. *et al.* Possible thermochemical disequilibrium in the atmosphere of the exoplanet GJ 436b. *Nature* **464**, 1161–1164 (2010).
- Stevenson, K. B. *et al.* Two nearby sub-Earth-sized exoplanet candidates in the GJ 436b system. *Astrophys. J.* **755**, 9 (2012).
- Knutson, H. A. *et al.* A *Spitzer* transmission spectrum for the exoplanet GJ 436b, evidence for stellar variability, and constraints on dayside flux variations. *Astrophys. J.* **735**, 27 (2011).
- Madhusudhan, N. & Seager, S. High metallicity and non-equilibrium chemistry in the dayside atmosphere of hot-Neptune GJ 436b. *Astrophys. J.* **729**, 41 (2011).
- Line, M. R. *et al.* Thermochemical and photochemical kinetics in cooler hydrogen-dominated extrasolar planets: a methane-poor GJ 436b? *Astrophys. J.* **738**, 32 (2011).
- Moses, J. I. *et al.* Compositional diversity in the atmospheres of hot Neptunes, with application to GJ 436b. *Astrophys. J.* **777**, 34 (2013).
- Figueira, P. *et al.* Bulk composition of the transiting hot Neptune around GJ 436. *Astron. Astrophys.* **493**, 671–676 (2009).
- Nettelmann, N. *et al.* Interior structure models of GJ 436b. *Astron. Astrophys.* **523**, A26 (2010).
- Miller, N. & Fortney, J. J. The heavy element masses of extrasolar giant planets, revealed. *Astrophys. J.* **736**, L29 (2011).
- McCullough, P. M. & MacKenty, J. *Considerations for using Spatial Scans with WFC3*. Instrument Science Report WFC3 2012–8 (Space Telescope Science Institute, 2012).
- Deming, D. *et al.* Infrared transmission spectroscopy of the exoplanets HD 209458b and XO-1b using the Wide Field Camera-3 on the Hubble Space Telescope. *Astrophys. J.* **774**, 95 (2013).
- Berta, Z. K. *et al.* The flat transmission spectrum of the super-Earth GJ 1214b from Wide Field Camera 3 on the Hubble Space Telescope. *Astrophys. J.* **747**, 35 (2012).
- Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J.* **580**, L171–L175 (2002).
- Gillon, M. *et al.* Improved parameters for the transiting hot Jupiters WASP-4b and WASP-5b. *Astron. Astrophys.* **496**, 259–267 (2009).
- Carter, J. A. & Winn, J. N. Parameter estimation from time-series data with correlated errors: a wavelet-based method and its application to transit light curves. *Astrophys. J.* **704**, 51–67 (2009).
- Sing, D. K. *et al.* Hubble Space Telescope transmission spectroscopy of the exoplanet HD 189733b: high-altitude atmospheric haze in the optical and near-ultraviolet with STIS. *Mon. Not. R. Astron. Soc.* **416**, 1443–1455 (2011).
- Pont, F. *et al.* The prevalence of dust on the exoplanet HD 189733b from Hubble and Spitzer observations. *Mon. Not. R. Astron. Soc.* **432**, 2917–2944 (2013).
- Benneke, B. & Seager, S. Atmospheric retrieval for super-Earth atmospheres: uniquely constraining the atmospheric composition with transmission spectroscopy. *Astrophys. J.* **753**, 100 (2012).
- Kreidberg, L. *et al.* Clouds in the atmosphere of the super-Earth exoplanet GJ 1214b. *Nature* <http://dx.doi.org/10.1038/nature12888> (this issue).
- Morley, C. V. *et al.* Quantitatively assessing the role of clouds in the transmission spectrum of GJ 1214b. *Astrophys. J.* **775**, 33 (2013).
- Adams, E. R., Seager, S. & Elkins-Tanton, L. Ocean planet or thick atmosphere: on the mass-radius relationship for solid exoplanets with massive atmospheres. *Astrophys. J.* **673**, 1160–1164 (2008).
- Baraffe, I., Chabrier, G. & Barman, T. Structure and evolution of super-Earth to super-Jupiter exoplanets. I. Heavy element enrichment in the interior. *Astron. Astrophys.* **482**, 315–332 (2008).
- Rogers, L. A. & Seager, S. A framework for quantifying the degeneracies of exoplanet interior compositions. *Astrophys. J.* **712**, 974–991 (2010).
- Ehrenreich, D., Lecavelier des Etangs, A. & Delfosse, X. HST/STIS Lyman- $\alpha$  observations of the quiet M dwarf GJ 436. *Astron. Astrophys.* **529**, A80 (2011).
- Heng, K. & Koppa, P. On the stability of super-Earth atmospheres. *Astrophys. J.* **754**, 60 (2012).
- Owen, J. E. & Jackson, A. P. Planetary evaporation by UV and X-ray radiation: basic hydrodynamics. *Mon. Not. R. Astron. Soc.* **425**, 2931–2947 (2012).
- Sanz-Forcada, J. *et al.* Estimation of the XUV radiation onto close planets and their evaporation. *Astron. Astrophys.* **532**, A6 (2011).
- Ehrenreich, D. & Désert, J.-M. Mass-loss rates for transiting exoplanets. *Astron. Astrophys.* **529**, A136 (2011).
- Fortney, J. J. *et al.* A framework for characterizing the atmospheres of low-mass low-density transiting planets. *Astrophys. J.* **775**, 80 (2013).
- Benneke, B. & Seager, S. How to distinguish between cloudy mini-Neptunes and water/volatile-dominated super-Earths. *Astrophys. J.* **778**, 153 (2013).

**Acknowledgements** We thank P. McCullough for his assistance in the planning and executing of these observations. We are also grateful to J. Moses, M. Line and N. Nettelmann for conversations on the nature of high-metallicity atmospheres as well as discussions of specific interior and atmosphere models for GJ 436b. D.H. has received support from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013 grant agreement number 247060).

**Author Contributions** H.A.K. carried out the data analysis for this project with input from D.D. B.B. provided the planetary atmosphere models and accompanying fits, and D.H. supplied the PHOENIX atmosphere models used to calculate the stellar limb-darkening coefficients.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.A.K. ([hknutson@caltech.edu](mailto:hknutson@caltech.edu)).

## METHODS

**Spectral extraction.** We follow the method described in a previous study<sup>11</sup> and summarize the main steps here for reference. We obtain our data using the  $256 \times 256$  pixel subarray and the SPARS10 mode with two samples (see section 13.3.6 of the HST Phase II Proposal Instructions for Cycle 21, available at <http://www.stsci.edu/ftp/documents/p2pi/p2pi.pdf>), which has an effective integration time of 7.62 s. These data are available for download from the Barbara A. Mikulski Archive for Space Telescopes (MAST) archive as part of proposal number 11622. We extract our spectra from the raw `_ima.fits` image files (see Extended Data Fig. 1 for a representative image), because the method for calculating the fluxes for the `_flt.fits` files does not work for data obtained in drift scan mode. The `_ima.fits` files were processed using either version 2.7 (all 2012 transits) or version 3.0 (Jan. 2013 transit) of the CALWF3 pipeline, which applies a standard set of calibrations including dark subtraction, linearity correction, cosmic ray rejection, and a conversion from raw counts to flux units as described in section 3.2.3 of the WFC3 Data Handbook.

The `_ima.fits` files retrieved through MAST contain an array of three images produced by the sample-up-the-ramp readout. These images were taken 0 s, 0.28 s and 7.62 s after the start of the exposure, and we refer to them as images A, B and C, respectively. We convert each image from units of electrons per second to electrons using the appropriate integration times and difference each pair of sampled images ( $B - A$  and  $C - B$ ). We then trim out a central region encompassing the location of the spectrum in each differenced image and add the trimmed differenced images together (that is,  $(B - A) + (C - B)$ ) to create our final science image (the working image that we use to extract the spectrum). Because we use a different sub-aperture for each differenced image, our final combined  $(B - A) + (C - B)$  science image is not simply equal to the  $(C - A)$  image. Previous studies<sup>11</sup> used this differenced image approach to minimize contamination from the sky background in the scanned images, therefore avoiding the need for a separate background subtraction step. Although we adopt the same approach in this study, our mask excludes only the sky background from the first 0.28 s of the 7.62 s integration and we therefore include a separate sky-subtraction step later in our analysis. We find that for our data this image differencing approach gives results that are identical to the case where we simply use the third (C) sample-up-the-ramp image as our science image with no subtraction or masking.

We next select a sub-aperture centred on the position of the stellar spectrum in our science image with dimensions of 160 pixels in the  $x$  (dispersion) direction and 71 pixels in the  $y$  (cross-dispersion) direction. We use a fixed aperture position for each visit, and estimate the position of the star using an acquisition image obtained at the start of each visit. Unlike previous studies<sup>11</sup>, we find that using a narrow aperture that cuts off at half the maximum flux produces an increased scatter in our white-light photometry; this may be due to the larger orbit-to-orbit position drift in our images compared to HD 204958b. In this case we obtain optimal results with an aperture that extends out to the wings of the point spread function in both dimensions.

We apply a colour-dependent flat-field correction and calculate wavelength solutions for our differenced images using coefficients adapted from the standard Space Telescope Science Institute pipeline as described in other studies (ref. 11 and Wilkins, A. N. *et al.*, personal communication). We then apply a filter to remove bad pixels and cosmic ray hits by first dividing each row of the individual subarrays by the total flux in that row (this corrects for the uneven scan rate in the  $y$  direction) and then iteratively flagging 8 $\sigma$  and then 4 $\sigma$  outliers in the time series at each pixel position using a moving median filter with a width of five pixels (that is, we calculate the median flux value at that pixel position starting from two images before and ending two images after our science image). We replace flagged pixels with the value of the moving median filter at that position, then multiply each row by the original flux total from that row to restore the initial subarray with bad pixels removed. Approximately 0.04%–0.06% of the pixels within the subarray aperture are flagged as bad in our four visits. We then sum in the  $y$  (cross-dispersion) direction to create a one-dimensional spectrum from each image.

We calculate the MJD mid-exposure times corresponding to each spectrum from the headers of the `.flt` files, and convert these times to BJD<sub>TDB</sub> using publicly available routines<sup>31</sup>. The median sky background in our `_ima.fits` images is approximately 0.1% of the total flux when we sum over the spectrum. We see no evidence for any wavelength or time dependence in the background flux, and so simply subtract the median background level from each visit.

As noted in a previous study<sup>11</sup>, the WFC3 spectra are undersampled and this can create problems when fitting templates with slightly offset positions in the dispersion direction. We mitigate this issue by convolving all of our one-dimensional spectra with a Gaussian function with a full width at half maximum (FWHM) of four pixels; this slightly degrades the wavelength resolution of our spectra, but the loss is negligible because we ultimately bin our transmission spectrum in four-pixel-wide bins. We next create a template spectrum for each visit by averaging ten spectra immediately before and immediately after the transit. We fit the template

spectrum to the central 112 pixels of the individual spectra from each visit, allowing the template amplitude to vary freely and the relative positions to shift in increments of a thousandth of a pixel.

**Transit fits.** The template-fitting technique results in two kinds of data product: first, a white-light curve for each transit calculated from the best-fit amplitude for the template spectrum at each time step, and second, a set of wavelength-dependent time series calculated from the difference of the best-fit template spectra and the measured spectra at each pixel position. This method is designed to remove common-mode white-light instrumental effects from the differenced spectra without the need to assume a functional form for these effects, resulting in lower noise levels in the final transmission spectrum than other commonly used approaches<sup>12,32,33</sup>.

Our fits to the white-light transit photometry (Fig. 1 and Extended Data Figs 2 and 3) include a linear function of time and a linear-plus-exponential function of orbital phase (five free parameters) to describe the behaviour of the instrument. We assume that  $i$  and  $a/R_*$  are the same in all visits, but allow  $R_p/R_*$ ,  $T_c$  and the instrumental terms to vary individually. We trim the data from the first orbit in our fits to the white-light data, because these data display larger-than-usual instrumental effects owing to settling at the new pointing. We keep this first orbit when we fit the differential transmission spectra (Extended Data Fig. 4) because there does not appear to be any evidence for colour-dependent instrumental effects at the start of each light curve and this gives us a longer baseline for our residual permutation error estimation. We set the uncertainties on each white-light measurement equal to the standard deviations of the residuals from our best-fit solution for that visit, which are  $10.0 \times 10^{-5}$  for the October,  $9.1 \times 10^{-5}$  for the November,  $8.9 \times 10^{-5}$  for the December and  $7.8 \times 10^{-5}$  for the January visits, respectively. These residuals are a factor of 1.2–1.5 times higher than the white-light photon noise limit of  $6.4 \times 10^{-5}$ , reflecting the uncorrected instrumental effects visible in Extended Data Fig. 3. The  $\chi^2$  value for our simultaneous fit to all four white-light transit curves is 356.7, with a total of 360 points in our fit and 26 free parameters. We also compare the root mean square of the residuals in our four-pixel bands to the photon noise limit in those bands and find median values ranging between 1.03–1.07 times the photon noise limit for our four individual transit observations. We calculate our errors on the wavelength-dependent transit depths.

We show our best-fit transit times in comparison to previously published values in Extended Data Fig. 5. Although we do not expect the planet radius to vary in time, previous studies have reported variations in the measured transit depths at different epochs, which could be caused by the occultation of bright or dark regions on the stellar surface<sup>33,34</sup>. We show the best-fit transit depths from our four white-light curves in comparison to these previous studies in Extended Data Fig. 6. In Extended Data Fig. 7 we plot the Ca II H and K activity index for this star as a function of time; although sampling is poor at the epoch of our HST observations, the star appears to have an average-to-low activity level at this time. This may explain the relatively small scatter in our measured transit depths over the two months spanned by our four transit observations compared to previous Spitzer observations. We calculate a reduced  $\chi^2$  value of 2.7 for our four transit depths compared to the averaged value, suggesting that stellar activity may still be contributing some extra variability.

For the wavelength-dependent transit fit, we calculate the transmission spectrum as follows: first we determine the difference between our extracted spectrum and a best-fit template spectrum at each pixel position and create a time series of the residuals. We then fit this time series with a model consisting of the difference between the white-light transit curve and a transit light curve with a freely varying planet–star radius ratio. We also include a linear function of time to account for the first order of any remaining instrumental trends. We compare the errors on the planet–star radius ratio from the Levenberg–Marquardt covariance matrix and the residual permutation method and take the larger of the two as our final uncertainty; they typically agree to within 10%. We then average the planet–star radius ratios in four-pixel-wide segments to create our final transmission spectrum for each visit (Extended Data Fig. 4), in which we select our wavelength range to exclude the low-illumination regions at the edges of the spectrum. We calculate uncertainties on each bin as the average of the errors for the four individual radius ratios to account for the four-pixel-wide Gaussian smoothing function we applied to the raw spectra before fitting the template spectrum. We combine the data from our four visits by taking the error-weighted mean of the transit depths in each wavelength bin (Fig. 2).

**Limb-darkening models.** We compare results for both our white-light fits and our differential transmission spectra using fixed four-parameter nonlinear limb-darkening coefficients calculated from a PHOENIX (<http://phoenix.astro.physik.uni-goettingen.de/>) stellar atmosphere model<sup>35</sup>. We first calculate the average centre-to-limb intensity profile for the nominal wavelength range of each individual pixel, then convolve the resulting model spectrum at each radial position on the star with a four-pixel-wide Gaussian function in wavelength space to account for the



smoothing applied to our measured spectra. We then fit the smoothed intensity profiles at each pixel position with a four-parameter nonlinear limb-darkening profile<sup>36</sup> and use those limb-darkening coefficients to calculate our transit light curves.

A recent study<sup>37</sup> estimated an effective temperature of  $3,416 \pm 54$  K for GJ 436 based on new interferometric radius measurements. We therefore consider four different stellar atmosphere models with effective temperatures ranging between 3,350 K and 3,500 K. We show plots of the disk-integrated spectra for the hottest and coldest models compared to the measured spectra for each visit in the WFC3 band in Extended Data Fig. 8. Our choice of limb-darkening model has a relatively small effect on the overall shape of our measured transmission spectrum, and we quantify this effect as a systematic error term in Extended Data Table 1. We estimate the contribution of the limb-darkening model errors by calculating the change in the measured transit depth in a given band over a range of 3,350–3,450 K in the stellar effective temperature used for the limb-darkening models. We then add these errors in quadrature to the measurement errors when comparing our results to model transmission spectra, and show the combined errors in Fig. 2.

We also compare the fits to the white-light curves using different stellar atmosphere models and find a  $\chi^2$  value of 357.2 for the 3,350 K model, 356.7 for the 3,400 K model, 357.2 for the 3,450 K model and 357.6 for the 3,500 K model, an effectively negligible difference. Without a strong preference for one model over the other, we elect to use the 3,400 K model in our final analysis for consistency with the published temperature estimate<sup>37</sup>. We tried fits with a linear limb-darkening coefficient as a free parameter at each wavelength, where we constrained these coefficients to vary within the range spanned by the model coefficients for stellar effective temperatures between 3,350 K and 3,450 K. We obtained a transmission spectrum that was consistent with our previous results, but with significantly larger uncertainties. This may be due to our choice of a linear parameterization for limb darkening, which provides a quantifiably poorer fit to the white-light transit curves, or to weak constraints on the limb-darkening profile due to GJ 436b's near-grazing geometry (impact parameter  $b = 0.85$ ; see our previous study of this planet<sup>3</sup> for a more detailed discussion of this geometry and its effect on our ability to empirically constrain limb-darkening profiles).

**Atmospheric retrieval.** The observed transmission spectrum is interpreted using a variant of the atmospheric retrieval method described in previous studies<sup>18,30</sup>. The method used in this work combines a self-consistent, line-by-line atmospheric forward model and the nested sampling technique to efficiently compute the joint posterior probability distribution of the desired atmospheric parameters from the observed transmission spectrum. The main variation from the method described in our most recent paper<sup>30</sup> is that the analysis in this work employs our a priori knowledge of chemistry to limit the range of atmospheric compositions to scenarios that are chemically plausible.

The goal of the retrieval analysis is to determine the range of metallicities (Fe/H) and cloud top pressures that are in agreement with the data (Fig. 3). Rather than fitting the data with unconstrained combinations of molecular abundances, however, we compare the observations only to atmospheres that are chemically plausible. Our approach is to compute the chemical equilibrium abundances and the temperature–pressure profiles self-consistently, while accounting for the uncertainties in the modelling of the methane abundance and the unknown Bond albedo through treating them as additional free parameters and marginalizing over them. In total, we perform a retrieval analysis in the five-dimensional-parameter space spanned by the metallicity, the cloud top pressure, the methane abundance relative to chemical equilibrium, the Bond albedo and the reference planet-to-star radius ratio.

We introduce a free parameter for the methane abundance because the methane abundance has a substantial effect on the observed part of the transmission spectrum, but its abundance profile cannot be predicted reliably using self-consistent models. The dominant source of uncertainty in our estimates for the methane abundance for a given metallicity is introduced by our limited knowledge of the vertical pressure–temperature profile. The proximity of the expected temperature profile to the CH<sub>4</sub>/CO transition<sup>6</sup> makes the methane abundance highly sensitive to the model assumptions about the vertical distribution of short-wavelength absorbers, vertical energy transport and day–night heat redistribution. Depending on whether the temperature in the photosphere is above or below the boundary at which CO replaces CH<sub>4</sub> as the dominant carbon-bearing species, the methane

abundance can vary by several orders of magnitude. Disequilibrium effects such as quenching and photochemistry at the upper end of the photosphere present an additional source of uncertainty. Our model determines the chemical composition of methane-reduced atmospheres by minimizing Gibb's free energy while simultaneously setting an upper limit on the methane abundance.

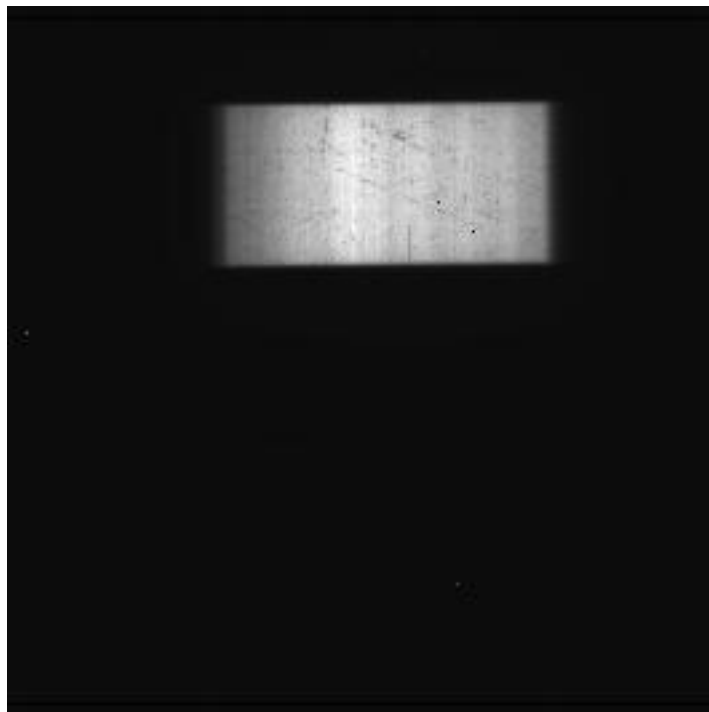
The other prominent absorber in the spectral range covered by our observations is water. We do not introduce an additional free parameter for water, however, because the water abundance in the photosphere can reliably be related to the metallicity of the atmosphere through chemical equilibrium calculations. Disequilibrium chemistry models for this planet<sup>6</sup> indicate that quenching and photochemical effects affect the water abundance only at pressures less than about 1  $\mu$ bar, whereas our data are primarily sensitive to higher pressures. We include the Bond albedo as a free parameter to account for the uncertainty introduced by the effect of the albedo on the temperature–pressure profile and therefore on the atmospheric scale height. When calculating the significance with which the solar-metallicity, cloud-free model is excluded we use the following definition<sup>38</sup>:

$$\text{Significance} = \frac{\chi_{\text{obs}}^2 - \langle \chi^2 \rangle}{\sigma} = \frac{\chi_{\text{obs}}^2 - \nu}{\sqrt{2\nu}}$$

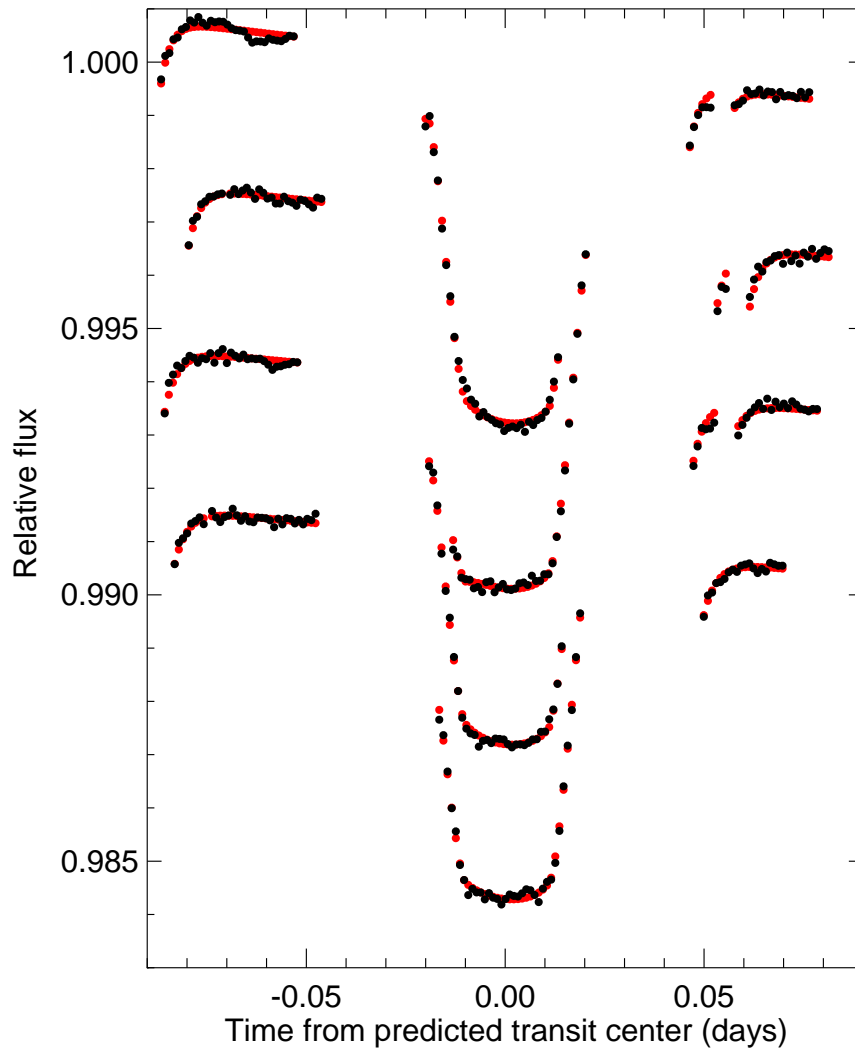
where  $\nu$  is the number of degrees of freedom in the fit. We calculate a value of  $48\sigma$  from our final fits.

We also consider scenarios with either subsolar (C/O = 0.3) or super-solar (C/O = 1.0) C to O ratios. We present the results from these retrievals in Extended Data Fig. 9. The oxygen-rich subsolar C to O case produces results that are virtually identical to our solar C to O analysis in Fig. 3. For the carbon-rich supersolar C to O case, the high metallicity ( $>1,000 \times$  solar) cloud-free scenarios are excluded because they exhibit a CO absorption feature at 1.6  $\mu$ m that appears to be inconsistent with our measured transmission spectrum. The same carbon-rich models also appear to allow a deeper ( $10 \times$  higher pressure) cloud deck for moderately metal-rich, several-hundred-times-solar scenarios. However, carbon-rich models with C/O  $> 0.8$  do not provide a good fit to GJ 436b's dayside emission spectrum, because increasing the C to O ratio tends to increase the amount of methane in the atmosphere<sup>6</sup>.

31. Eastman, J., Siverd, R. & Gaudi, B. S. Achieving better than 1 minute accuracy in the heliocentric and barycentric Julian dates. *Proc. Astron. Soc. Pacific* **122**, 935–946 (2010).
32. Knutson, H. A. *et al.* Using stellar limb-darkening to refine the properties of HD 209458b. *Astrophys. J.* **655**, 564 (2007).
33. Swain, M. R., Vasisht, G. & Tinetti, G. The presence of methane in the atmosphere of an extrasolar planet. *Nature* **452**, 329–331 (2008).
34. Beaulieu, J. P. *et al.* Methane in the atmosphere of the transiting hot Neptune GJ 436b? *Astrophys. J.* **731**, 16 (2011).
35. Hauschildt, P. *et al.* The NEXTGEN Model Atmosphere Grid. II. Spherically symmetric model atmospheres for giant stars with effective temperatures between 3000–6800 K. *Astrophys. J.* **525**, 871–880 (1999).
36. Claret, A. A new non-linear limb-darkening law for LTE stellar atmosphere models. *Astron. Astrophys.* **363**, 1081–1090 (2000).
37. von Braun, K. *et al.* The GJ 436 system: directly determined astrophysical parameters of an M dwarf and implications for the transiting hot Neptune. *Astrophys. J.* **753**, 171 (2012).
38. Gregory, P. *Bayesian Logical Data Analysis for the Physical Sciences* 167 (Cambridge Univ. Press, 2005).
39. Bean, J. L. *et al.* A Hubble Space Telescope transit light curve for GJ 436b. *Astron. Astrophys.* **486**, 1039–1046 (2008).
40. Coughlin, J. L. *et al.* New observations and a possible detection of parameter variations in the transits of Gliese 436b. *Astrophys. J.* **689**, L149–L152 (2008).
41. Alonso, R. *et al.* Limits to the planet candidate GJ 436c. *Astron. Astrophys.* **487**, L5–L8 (2008).
42. Pont, F. *et al.* Transit infrared spectroscopy of the hot Neptune around GJ 436 with the Hubble Space Telescope. *Mon. Not. R. Astron. Soc.* **393**, L6–L10 (2009).
43. Shporer, A. *et al.* Photometric follow-up observations of the transiting Neptune-mass planet GJ 436b. *Astrophys. J.* **694**, 1559–1565 (2009).
44. Cáceres, C. *et al.* High cadence near infrared timing observations of extrasolar planets I. GJ 436b and XO-1b. *Astron. Astrophys.* **507**, 481–486 (2009).
45. Ballard, S. *et al.* A search for additional planets in the NASA EPOXI observations of the exoplanet system GJ 436. *Astrophys. J.* **716**, 1047–1059 (2010).
46. Knutson, H. A., Howard, A. W. & Isaacson, H. A correlation between stellar activity and hot Jupiter emission spectra. *Astrophys. J.* **720**, 1569–1576 (2010).



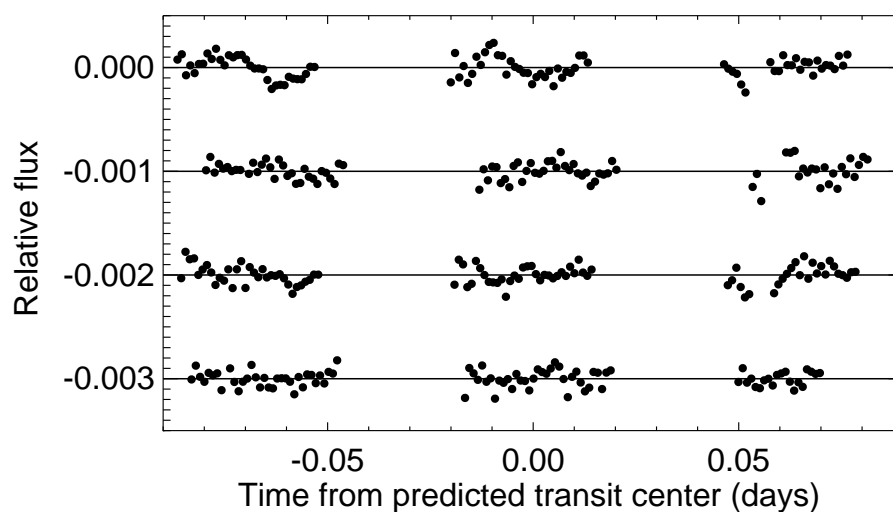
Extended Data Figure 1 | Representative raw image from the 29 November 2012 UT observation, showing the scanned spectrum.



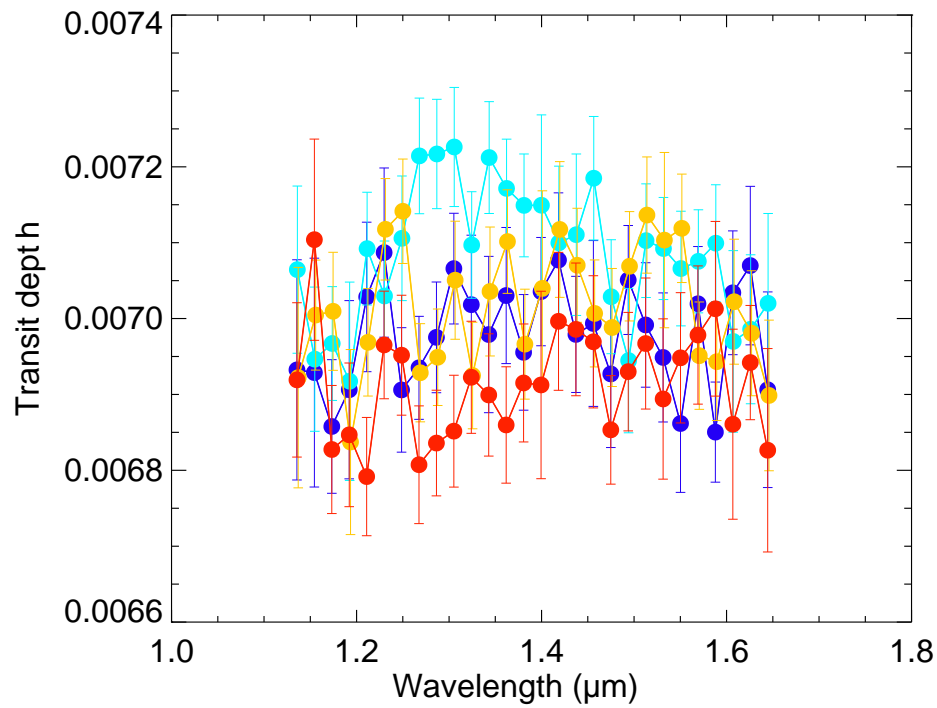
**Extended Data Figure 2 | Raw white-light photometry for the four individual transits.** Data are vertically offset for clarity. Transits shown were obtained on the following dates (from top to bottom): 26 October 2012 UT,

29 November 2012 UT, 10 December 2012 UT and 2 January 2013 UT. The raw fluxes are shown as filled black circles, and the best-fit solutions for the instrumental effects and transit light curves are shown as filled red circles.



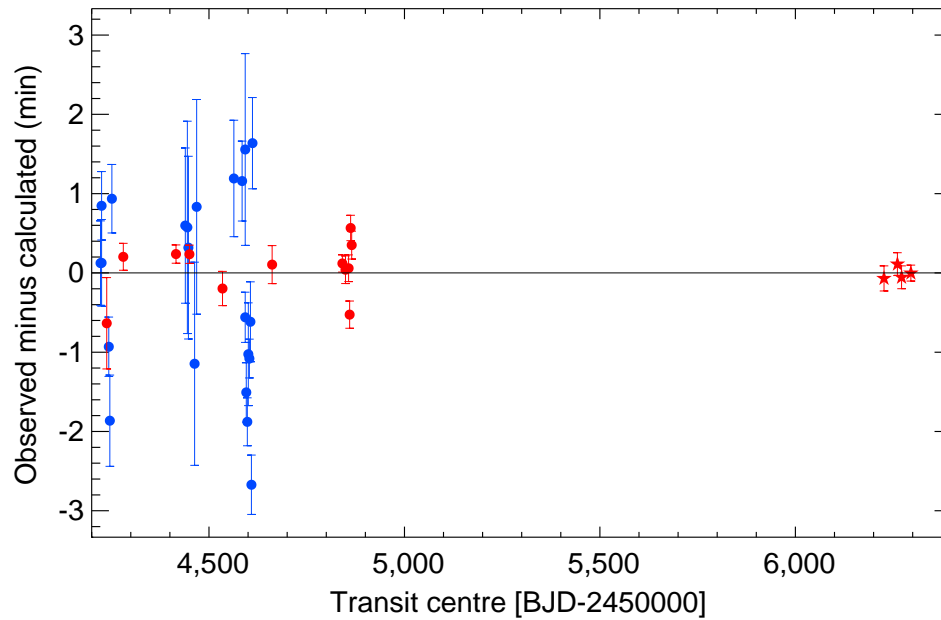


**Extended Data Figure 3 | White-light residuals.** Data are vertically offset for clarity. Transit residuals shown were obtained on the following dates (from top to bottom): 26 October 2012 UT, 29 November 2012 UT, 10 December 2012 UT and 2 January 2013 UT. The difference between the white-light fluxes and best-fit model solutions are shown as filled black circles.



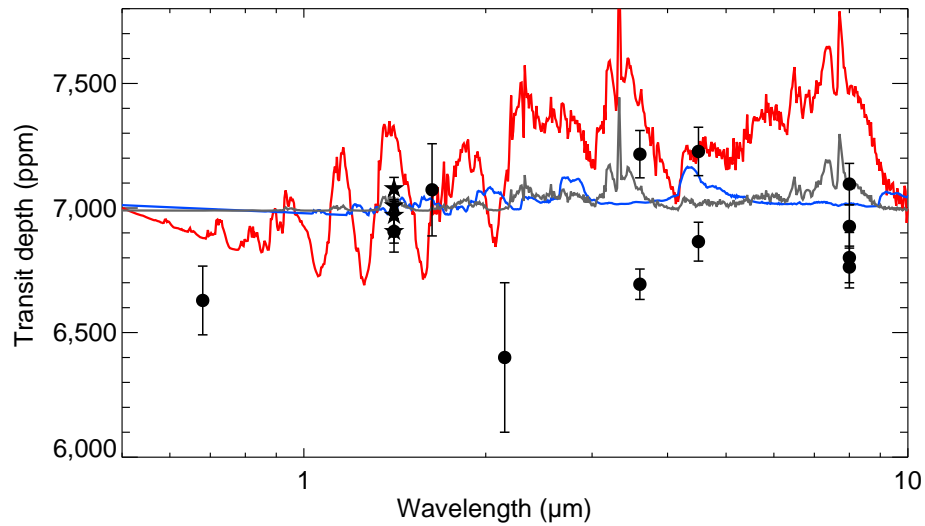
**Extended Data Figure 4 | Individual transmission spectra for each of the four visits.** Transmission spectra are shown as filled circles, with colours indicating the date of the observations: 26 October 2012 UT (dark blue), 29 November 2012 UT (light blue), 10 December 2012 UT (yellow) and

2 January 2013 UT (red). This plot shows the errors in the measured transit depths, but does not include the additional systematic errors from the limb-darkening models.



**Extended Data Figure 5 | Observed minus calculated transit times using the new best-fit ephemeris.** The solid line denotes observed minus calculated equal to zero. Transit times from this paper are plotted as filled stars, and previously published observations are shown as filled circles, with  $1\sigma$

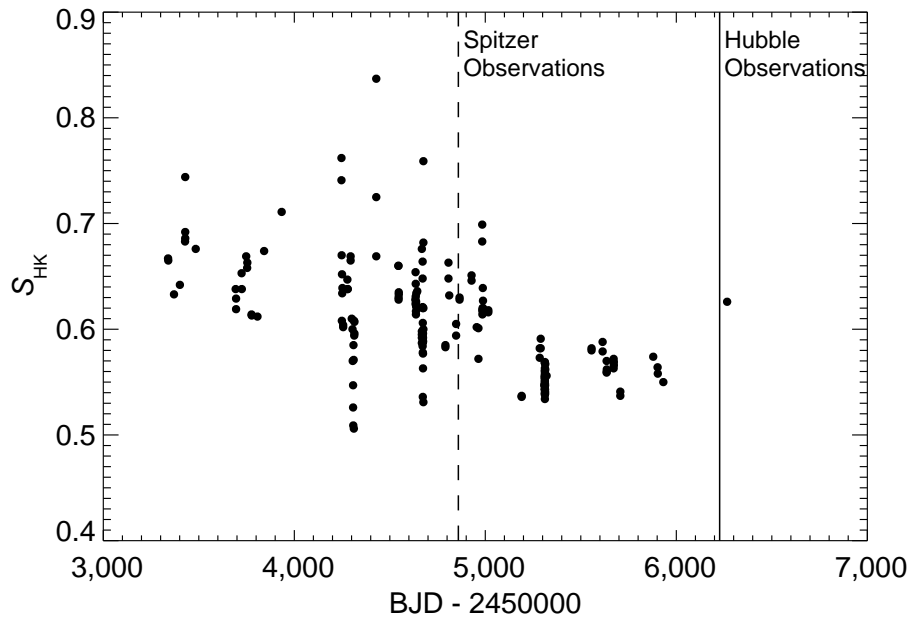
uncertainties overplotted for both. The colour of the points denotes the wavelength of the observations (blue for visible, red for infrared). Transits shown include all previously published observations for this planet<sup>3,39–45</sup>. The figure is adapted from figure 8 of our previous study<sup>3</sup>.



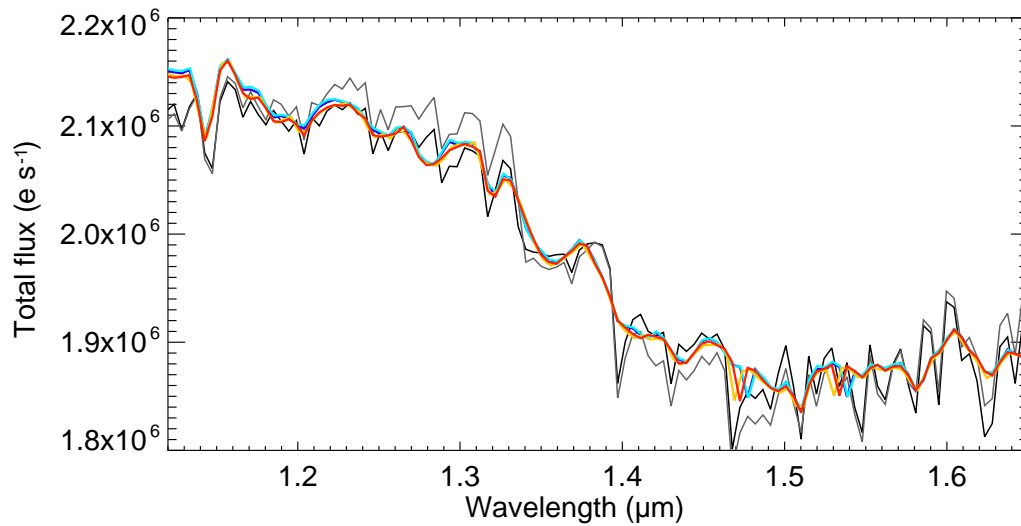
**Extended Data Figure 6 | Comparison to published transit depths for GJ 436b.** Filled black circles show previously published transit depths<sup>3,42,43,45,46</sup>, with  $1\sigma$  uncertainties overplotted. The white-light transit depths from our WFC3 observations are plotted as black stars. We show three models for comparison, including a solar-metallicity cloud-free model (red line), a hydrogen-poor  $1,900 \times$  solar model (blue line), and a solar metallicity model

with optically thick clouds at 1 mbar (grey line). As we discuss in ref. 3, the apparent variations in transit depth at different epochs could plausibly be explained by the occultation of active regions on the surface of the star. If correct, this would make broadband photometry collected at different epochs unreliable for the purpose of constraining the planet's transmission spectrum. p.p.m., parts per million.



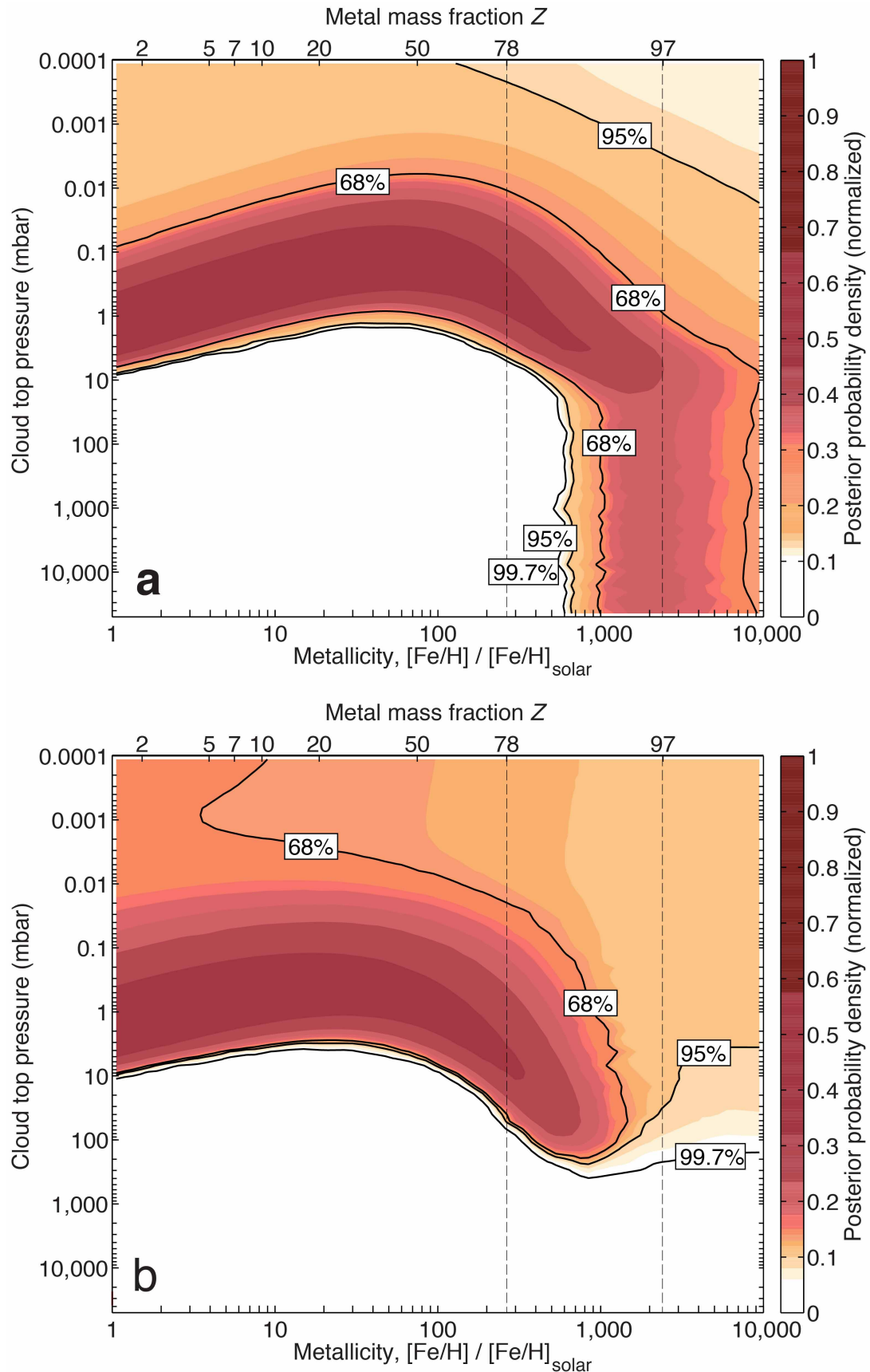


**Extended Data Figure 7 | Stellar activity versus time.** Filled black circles show the measured emission levels in the Ca II H and K line cores from Keck HIgh Resolution Echelle Spectrometer (HIRES) spectroscopy of GJ 436 (refs 3 and 46); these are parameterized using the  $S_{\text{HK}}$  index, where larger values indicate increased stellar activity. Vertical lines mark the approximate dates of the six most recent Spitzer transit observations (dashed line), as well as the four HST transits presented in this paper (solid line).



**Extended Data Figure 8 | Averaged stellar spectrum versus PHOENIX model atmospheres.** Spectra are averaged over each HST visit and then normalized using the sensitivity curve for that visit. These spectra are plotted as dark blue (October), light blue (November), yellow (December), and red (January) lines. For comparison we show two PHOENIX stellar atmosphere

models with effective temperatures of 3,500 K and  $\log(g) = 5.0$  (black line, where  $g$  is surface gravity) and 3,350 K and  $\log(g) = 4.8$  (grey line) binned to the same pixel resolution as our data. These data include an additional component of instrumental broadening that smoothes out the sharp spectral features visible in the model spectra.



**Extended Data Figure 9 | Joint constraints on cloud-top pressure versus atmospheric metallicity for an oxygen-rich and a carbon-rich atmosphere.** a,  $\text{C}/\text{O} = 0.3$ ; b,  $\text{C}/\text{O} = 1.0$ . The models shown in Fig. 3 assume a solar  $\text{C}/\text{O}$  ratio of 0.5. The coloured shading indicates the normalized probability density as a function

of cloud-top pressure and metallicity derived from a variation of the Bayesian retrieval methods. We vary the amount of metals in the atmosphere (defined as elements heavier than H and He) linearly using the scaling factor shown on the lower  $x$  axes. The black contours show the 68%, 95% and 99.7% Bayesian credible regions.

Extended Data Table 1 | Averaged differential transit depths

Wavelength	Depth (p.p.m.)	Measurement error (p.p.m.)	Error from limb darkening (p.p.m.)
1.136	6966	60	7
1.155	6994	50	10
1.174	6924	40	12
1.193	6872	57	12
1.211	6968	39	17
1.230	7046	38	22
1.249	7036	39	20
1.268	6967	35	22
1.289	6989	35	24
1.306	7043	38	17
1.324	6989	38	20
1.343	7046	42	19
1.362	7057	37	25
1.381	7006	37	25
1.400	7036	50	27
1.419	7072	46	46
1.438	7030	42	46
1.456	7044	42	44
1.475	6948	39	44
1.494	7008	39	49
1.513	7057	40	55
1.532	7022	44	56
1.551	7018	40	46
1.570	7010	37	41
1.588	6959	40	41
1.607	6994	44	34
1.626	6984	44	30
1.645	6916	59	55



# Clouds in the atmosphere of the super-Earth exoplanet GJ 1214b

Laura Kreidberg<sup>1</sup>, Jacob L. Bean<sup>1</sup>, Jean-Michel Désert<sup>2,3</sup>, Björn Benneke<sup>4</sup>, Drake Deming<sup>5</sup>, Kevin B. Stevenson<sup>1</sup>, Sara Seager<sup>4</sup>, Zachory Berta-Thompson<sup>6,7</sup>, Andreas Seifahrt<sup>1</sup> & Derek Homeier<sup>8</sup>

Recent surveys have revealed that planets intermediate in size between Earth and Neptune ('super-Earths') are among the most common planets in the Galaxy<sup>1–3</sup>. Atmospheric studies are the next step towards developing a comprehensive understanding of this new class of object<sup>4–6</sup>. Much effort has been focused on using transmission spectroscopy to characterize the atmosphere of the super-Earth archetype GJ 1214b (refs 7–17), but previous observations did not have sufficient precision to distinguish between two interpretations for the atmosphere. The planet's atmosphere could be dominated by relatively heavy molecules, such as water (for example, a 100 per cent water vapour composition), or it could contain high-altitude clouds that obscure its lower layers. Here we report a measurement of the transmission spectrum of GJ 1214b at near-infrared wavelengths that definitively resolves this ambiguity. The data, obtained with the Hubble Space Telescope, are sufficiently precise to detect absorption features from a high mean-molecular-mass atmosphere. The observed spectrum, however, is featureless. We rule out cloud-free atmospheric models with compositions dominated by water, methane, carbon monoxide, nitrogen or carbon dioxide at greater than 5 $\sigma$  confidence. The planet's atmosphere must contain clouds to be consistent with the data.

We observed 15 transits of the planet GJ 1214b with the Wide Field Camera 3 (WFC3) instrument on the Hubble Space Telescope (HST) between 27 September 2012 and 22 August 2013 Universal Time (UT). Each transit observation (visit) consisted of four orbits of the telescope, with 45-min gaps in phase coverage between target visibility periods due to Earth occultation. We obtained time-series spectroscopy from 1.1  $\mu\text{m}$  to 1.7  $\mu\text{m}$  during each observation. The data were taken in spatial scan mode, which slews the telescope during the exposure and moves the spectrum perpendicularly to the dispersion direction on the detector. This mode reduces the instrumental overhead time by a factor of five compared to staring mode observations. We achieved an integration efficiency of 60%–70%. We extracted the spectra and divided each exposure into five-pixel-wide bins, obtaining spectrophotometric time series in 22 channels (resolution  $R \equiv \lambda/\Delta\lambda \approx 70$ ). The typical signal-to-noise ratio per 88.4-s exposure per channel was 1,400. We also created a 'white' light curve summed over the entire wavelength range. Our analysis incorporates data from 12 of the 15 transits observed, because one observation was compromised by a telescope guiding error and two showed evidence of a starspot crossing.

The raw transit light curves for GJ 1214b exhibit ramp-like systematics comparable to those seen in previous WFC3 data<sup>10,18,19</sup>. The ramp in the first orbit of each visit consistently has the largest amplitude and a different shape from ramps in the subsequent orbits. Following standard procedure for HST transit light curves, we did not include data from the first orbit in our analysis, leaving 654 exposures. We corrected for systematics in the remaining three orbits using two techniques that have been successfully applied in prior analyses<sup>10,18,20</sup>. The first approach models the systematics as an analytic function of time. The function

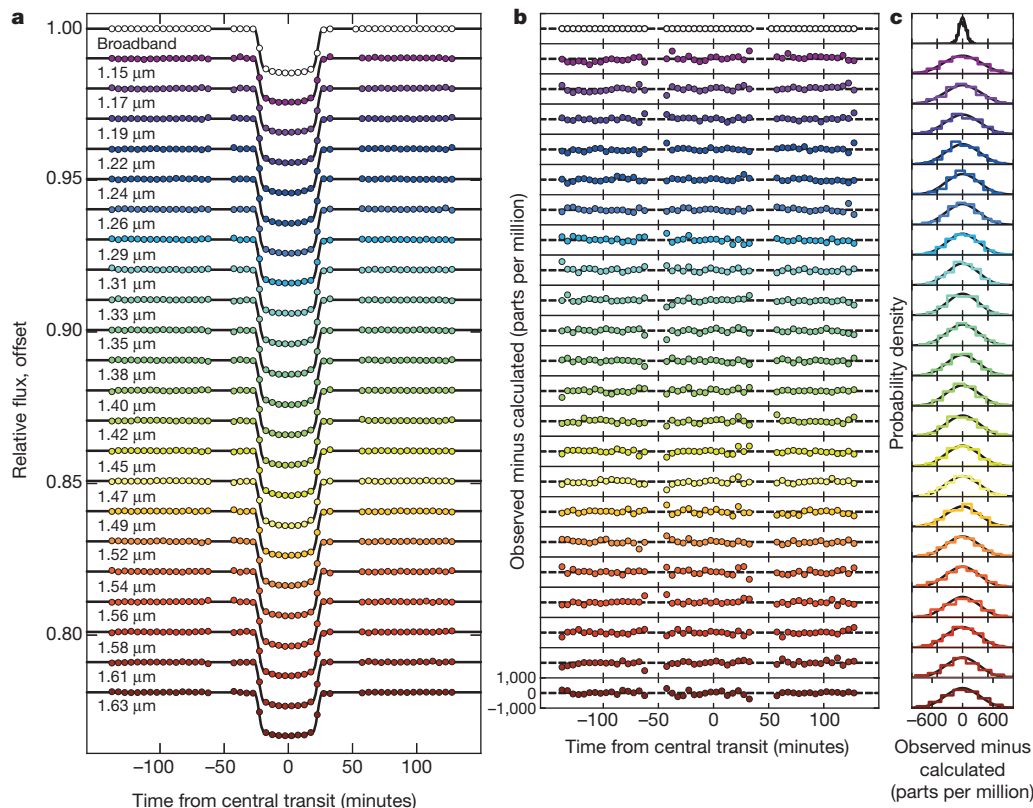
includes an exponential ramp term fitted to each orbit, a linear trend with time for each visit, and a normalization factor. The second approach assumes that the morphology of the systematics is independent of wavelength, and models each channel with a scalar multiple of the time series of systematics from the white-light curve fit. We obtained consistent results from both methods (see Extended Data Table 1), and report here results from the second. See the Supplementary Information and Extended Data Figs 1–6 for more detail on the observations, data reduction and systematics correction.

We fitted the light curves in each spectroscopic channel with a transit model<sup>21</sup> to measure the transit depth as a function of wavelength; this constitutes the transmission spectrum. See Fig. 1 for the fitted transit light curves. We used the second systematics correction technique described above and fitted a unique planet-to-star radius ratio  $R_p/R_s$  and normalization  $C$  to each channel and each visit, and a unique linear limb-darkening parameter  $u$  to each channel. We assumed a circular orbit<sup>22</sup> and fixed the inclination  $i$  to be 89.1°, the ratio of the semi-major axis to the stellar radius  $a/R_s$  to be 15.23, the orbital period  $P$  to be 1.58040464894 days, and the time of central transit  $T_c$  to be 2454966.52488 BJD<sub>TDB</sub>. These are the best-fit values to the white-light curve.

The measured transit depths in each channel are consistent over all transit epochs (see Extended Data Fig. 5), and we report the weighted average depth per channel. The resulting transmission spectrum is shown in Fig. 2. Our results are not significantly affected by stellar activity, as we discuss further in the Supplementary Information. Careful treatment of the limb darkening is critical to the results, but our limb-darkening measurements are not degenerate with the transit depth (see Extended Data Fig. 4) and agree with the predictions from theoretical models (see Extended Data Fig. 6). Our conclusions are unchanged if we fix the limb darkening on theoretical values. We find that a linear limb-darkening law is sufficient to model the data. For further description of the limb-darkening treatment, see the Supplementary Information.

The transmission spectrum we report here has the precision necessary to detect the spectral features of a high-mean-molecular-mass atmosphere for the first time. However, the observed spectrum is featureless. The data are best-fitted with a flat line, which has a reduced  $\chi^2$  of 1.0. We compare several models to the data that represent limiting-case scenarios in the range of expected atmospheric compositions<sup>17,23</sup>. Depending on the formation history and evolution of the planet, a high-mean-molecular-mass atmosphere could be dominated by water (H<sub>2</sub>O), methane (CH<sub>4</sub>), carbon monoxide (CO), carbon dioxide (CO<sub>2</sub>) or nitrogen (N<sub>2</sub>). Water is expected to be the dominant absorber in the wavelength range of our observations, so a wide range of high-mean-molecular-mass atmospheres with trace amounts of water can be approximated by a pure H<sub>2</sub>O model. The data show no evidence for water absorption. A cloud-free pure H<sub>2</sub>O composition is ruled out at 16.1 $\sigma$  confidence. In the case of a dry atmosphere, features from other absorbers such as CH<sub>4</sub>, CO or CO<sub>2</sub> could be visible in the transmission spectrum. Cloud-free

<sup>1</sup>Department of Astronomy and Astrophysics, University of Chicago, Chicago, Illinois 60637, USA. <sup>2</sup>CASA, Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, Colorado 80309, USA. <sup>3</sup>Department of Astronomy, California Institute of Technology, Pasadena, California 91101, USA. <sup>4</sup>Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>5</sup>Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. <sup>6</sup>Department of Astronomy, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>7</sup>MIT Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>8</sup>Centre de Recherche Astrophysique de Lyon, 69364 Lyon, France.



**Figure 1 | Spectrophotometric data for transit observations of GJ 1214b.** **a**, Normalized and systematics-corrected data (points) with best-fit transit models (lines), offset for clarity. The data consist of 12 transit observations and are binned in phase in 5-min increments. The spectroscopic light curve fit parameters are transit depth, a linear limb-darkening coefficient, and a normalization term to correct for systematics. A unique transit depth is determined for each observation and the measured transit depths are consistent from epoch to epoch in all channels. **b**, Binned residuals from the best-fit model light curves. The residuals are within

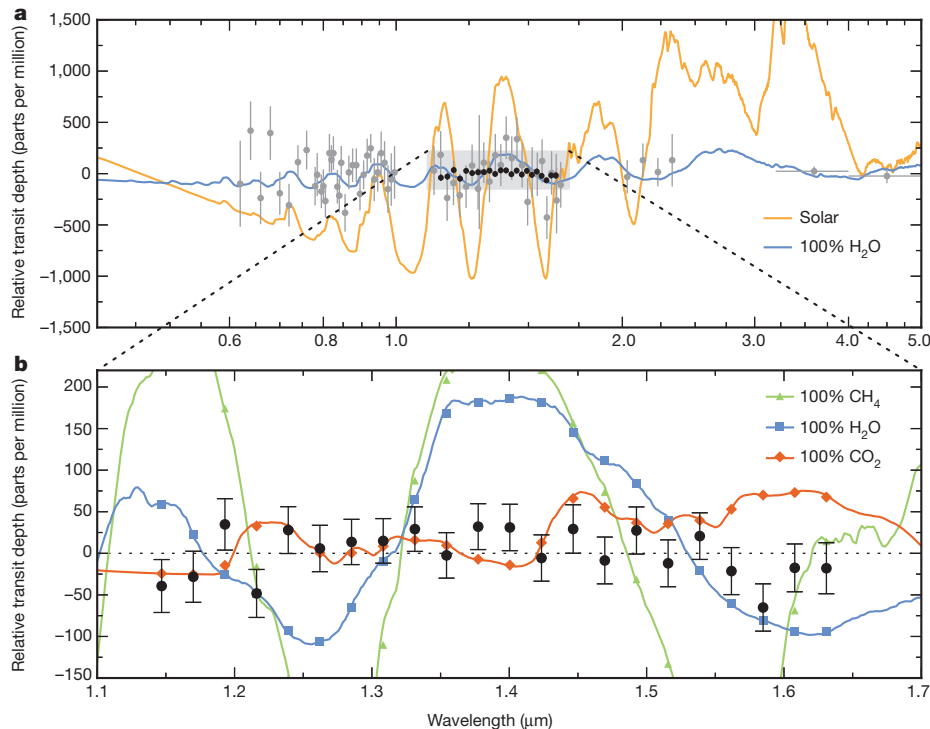
14% of the predicted photon-limited shot noise in all spectroscopic channels. The median observed root-mean-square in the spectroscopic channels is 315 p.p.m., before binning. **c**, Histograms of the unbinned residuals (coloured lines) compared to the expected photon noise (black lines). The residuals are Gaussian, satisfying a Shapiro–Wilk test for normality at the  $\alpha = 0.1$  level in all but one channel (1.24  $\mu\text{m}$ ). The median reduced  $\chi^2$  value for the spectroscopic light curve fits is 1.02.

atmospheres composed of these absorbers are also excluded by the data, at  $31.1\sigma$ ,  $7.5\sigma$  and  $5.5\sigma$  confidence, respectively. Nitrogen has no spectral features in the observed wavelength range, but our measurements are sensitive to a nitrogen-rich atmosphere with trace amounts of spectrally active molecules. For example, we can rule out a 99.9%  $\text{N}_2$ , 0.1%  $\text{H}_2\text{O}$  atmosphere at  $5.6\sigma$  confidence. Of the scenarios considered here, a 100%  $\text{CO}_2$  atmosphere is the most challenging to detect because  $\text{CO}_2$  has the highest molecular mass and a relatively small opacity in the observed wavelength range. Given that the data are precise enough to rule out even a  $\text{CO}_2$  composition at high confidence, the most likely explanation for the absence of spectral features is a grey opacity source, suggesting that clouds are present in the atmosphere. Clouds can block transmission of stellar flux through the atmosphere, which truncates spectral features arising from below the cloud altitude<sup>24</sup>.

To illustrate the properties of potential clouds, we perform a Bayesian analysis on the transmission spectrum with a code designed for spectral retrieval of super-Earth atmospheric compositions<sup>25</sup>. We assume a two-component model atmosphere of water and a solar mix of hydrogen and helium gas, motivated by the fact that water is the most abundant icy volatile for solar abundance ratios. Clouds are modelled as a grey, optically thick opacity source below a given altitude. See Fig. 3 for the retrieval results. For this model, the data constrain the cloud-top pressure to less than  $10^{-2}$  mbar for a mixing ratio with mean molecular mass equal to solar and less than  $10^{-1}$  mbar for a water-dominated composition (both at  $3\sigma$  confidence). At the temperatures and pressures expected in the atmosphere of GJ 1214b, equilibrium condensates of ZnS and KCl could form in the observable part of the atmosphere. Although these species could provide the necessary opacity, they are

predicted to form at much higher pressures (deeper than 10 mbar for a  $50\times$  solar metallicity model)<sup>16</sup>, requiring that clouds be lofted high from their base altitude to explain the spectrum we measured. Alternatively, photochemistry could produce a layer of hydrocarbons in the upper atmosphere, analogous to the haze on Saturn's moon Titan<sup>14,16</sup>.

The result presented here demonstrates the capability of current facilities to measure very precise spectra of exoplanets by combining many transit observations. This observational strategy has the potential to yield the atmospheric characterization of an Earth-size planet orbiting in the habitable zone of a small, nearby star. Transmission spectrum features probing five scale heights of a nitrogen-rich atmosphere on such a planet would have an amplitude of 30 parts per million (p.p.m.), which is comparable to the photon-limited measurement precision we obtained with the Hubble Space Telescope. However, our findings for the super-Earth archetype GJ 1214b, as well as emerging results for other exoplanets<sup>18,26,27</sup>, suggest that clouds may exist across a wide range of planetary atmosphere compositions, temperatures and pressures. Clouds do not generally have constant opacity at all wavelengths, so further progress in this area can be made by obtaining high-precision data with broad spectral coverage. Another way forward is to focus on measuring exoplanet emission and reflection spectra during secondary eclipse, because the optical depth of clouds viewed at near-normal incidence is lower than that for the slant geometry observed during transit<sup>24</sup>. Fortunately, the next generation of large ground-based telescopes and the James Webb Space Telescope will have the capability to make such measurements, bringing us within reach of characterizing potentially habitable worlds beyond our Solar System.

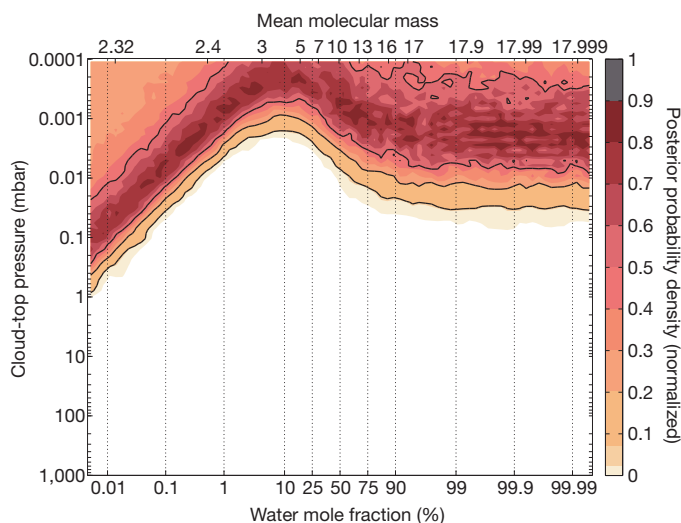


**Figure 2 | The transmission spectrum of GJ 1214b.** **a**, Transmission spectrum measurements from our data (black points) and previous work (grey points)<sup>7–11</sup>, compared to theoretical models (lines). The error bars correspond to  $1\sigma$  uncertainties. Each data set is plotted relative to its mean. Our measurements are consistent with past results for GJ 1214b using WFC3 (ref. 10). Previous data rule out a cloud-free solar composition (orange line), but are consistent with either a high-mean-molecular-mass atmosphere (for example, 100% water, blue line) or a hydrogen-rich atmosphere with high-altitude clouds. **b**, Detailed view of our measured transmission spectrum (black points) compared to high-mean-molecular-mass models (lines). The

error bars are  $1\sigma$  uncertainties in the posterior distribution from a Markov chain Monte Carlo fit to the light curves (see the Supplementary Information for details of the fits). The coloured points correspond to the models binned at the resolution of the observations. The data are consistent with a featureless spectrum ( $\chi^2 = 21.1$  for 21 degrees of freedom), but inconsistent with cloud-free high-mean-molecular-mass scenarios. Fits to pure water (blue line), methane (green line), carbon monoxide (not shown), and carbon dioxide (red line) models have  $\chi^2 = 334.7, 1067.0, 110.0$  and  $75.4$  with 21 degrees of freedom, and are ruled out at  $16.1\sigma, 31.1\sigma, 7.5\sigma$  and  $5.5\sigma$  confidence, respectively.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 August; accepted 18 November 2013.



**Figure 3 | Spectral retrieval results for a two-component (hydrogen/helium and water) model atmosphere for GJ 1214b.** The colours indicate posterior probability density as a function of water mole fraction and cloud-top pressure. Black contours mark the  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  Bayesian credible regions. Clouds are modelled as having a grey opacity, with transmission truncated below the cloud altitude. The atmospheric modelling assumes a surface gravity of  $8.48 \text{ m s}^{-2}$  and an equilibrium temperature equal to 580 K.

1. Cassan, A. *et al.* One or more bound planets per Milky Way star from microlensing observations. *Nature* **481**, 167–169 (2012).
2. Fressin, F. *et al.* The false positive rate of Kepler and the occurrence of planets. *Astrophys. J.* **766**, 81 (2013).
3. Petigura, E. A., Marcy, G. W. & Howard, A. W. A plateau in the planet population below twice the size of Earth. *Astrophys. J.* **770**, 69 (2013).
4. Adams, E. R., Seager, S. & Elkins-Tanton, L. Ocean planet or thick atmosphere: on the mass-radius relationship for solid exoplanets with massive atmospheres. *Astrophys. J.* **673**, 1160–1164 (2008).
5. Miller-Ricci, E., Seager, S. & Sasselov, D. The atmospheric signatures of super-Earths: how to distinguish between hydrogen-rich and hydrogen-poor atmospheres. *Astrophys. J.* **690**, 1056–1067 (2009).
6. Rogers, L. A. & Seager, S. Three possible origins for the gas layer on GJ 1214b. *Astrophys. J.* **716**, 1208–1216 (2010).
7. Bean, J. L., Miller-Ricci Kempton, E. & Homeier, D. A ground-based transmission spectrum of the super-Earth exoplanet GJ 1214b. *Nature* **468**, 669–672 (2010).
8. Désert, J.-M. *et al.* Observational evidence for a metal-rich atmosphere on the super-Earth GJ1214b. *Astrophys. J.* **731**, L40 (2011).
9. Bean, J. L. *et al.* The optical and near-infrared transmission spectrum of the super-Earth GJ 1214b: further evidence for a metal-rich atmosphere. *Astrophys. J.* **743**, 92 (2011).
10. Berta, Z. K. *et al.* The flat transmission spectrum of the super-Earth GJ1214b from Wide Field Camera 3 on the Hubble Space Telescope. *Astrophys. J.* **747**, 35 (2012).
11. Fraine, J. D. *et al.* Spitzer transits of the super-Earth GJ1214b and implications for its atmosphere. *Astrophys. J.* **765**, 127 (2013).
12. Miller-Ricci, E. & Fortney, J. J. The nature of the atmosphere of the transiting super-Earth GJ 1214b. *Astrophys. J.* **716**, L74–L79 (2010).
13. Nettelmann, N., Fortney, J. J., Kramm, U. & Redmer, R. Thermal evolution and structure models of the transiting super-Earth GJ 1214b. *Astrophys. J.* **733**, 2 (2011).
14. Miller-Ricci Kempton, E., Zahnle, K. & Fortney, J. J. The atmospheric chemistry of GJ 1214b: photochemistry and clouds. *Astrophys. J.* **745**, 3 (2012).

15. Howe, A. R. & Burrows, A. S. Theoretical transit spectra for GJ 1214b and other “super-Earths”. *Astrophys. J.* **756**, 176 (2012).
16. Morley, C. V. *et al.* Quantitatively assessing the role of clouds in the transmission spectrum of GJ 1214b. *Astrophys. J.* **775**, 33 (2013).
17. Benneke, B. & Seager, S. How to distinguish between cloudy mini-Neptunes and water/volatile-dominated super-Earths. *Astrophys. J.* **778**, 153 (2013).
18. Deming, D. *et al.* Infrared transmission spectroscopy of the exoplanets HD 209458b and XO-1b using the Wide Field Camera-3 on the Hubble Space Telescope. *Astrophys. J.* **774**, 95 (2013).
19. Swain, M. *et al.* Probing the extreme planetary atmosphere of WASP-12b. *Icarus* **225**, 432–445 (2013).
20. Stevenson, K. B. *et al.* Transmission spectroscopy of the hot-Jupiter WASP-12b from 0.7 to 5 microns. <http://arxiv.org/abs/1305.1670> (2013).
21. Mandel, K. & Agol, E. Analytic light curves for planetary transit searches. *Astrophys. J.* **580**, L171–L175 (2002).
22. Anglada-Escudé, G., Rojas-Ayala, B., Boss, A. P., Weinberger, A. J. & Lloyd, J. P. GJ 1214 reviewed. Trigonometric parallax, stellar parameters, new orbital solution, and bulk properties for the super-Earth GJ 1214b. *Astron. Astrophys.* **551**, A48 (2013).
23. Fortney, J. J. *et al.* A framework for characterizing the atmospheres of low-mass low-density transiting planets. *Astrophys. J.* **775**, 80 (2013).
24. Fortney, J. J. The effect of condensates on the characterization of transiting planet atmospheres with transmission spectroscopy. *Mon. Not. R. Astron. Soc.* **364**, 649–653 (2005).
25. Benneke, B. & Seager, S. Atmospheric retrieval for super-Earths: uniquely constraining the atmospheric composition with transmission spectroscopy. *Astrophys. J.* **753**, 100 (2012).
26. Pont, F., Knutson, H., Gilliland, R. L., Moutou, C. & Charbonneau, D. Detection of atmospheric haze on an extrasolar planet: the 0.55–1.05  $\mu\text{m}$  transmission spectrum of HD 189733b with the Hubble Space Telescope. *Mon. Not. R. Astron. Soc.* **385**, 109–118 (2008).
27. Knutson, H. A., Benneke, B., Deming, D. & Homeier, D. A featureless transmission spectrum for the Neptune-mass exoplanet GJ 436b. *Nature* <http://dx.doi.org/10.1038/nature12887> (this issue).

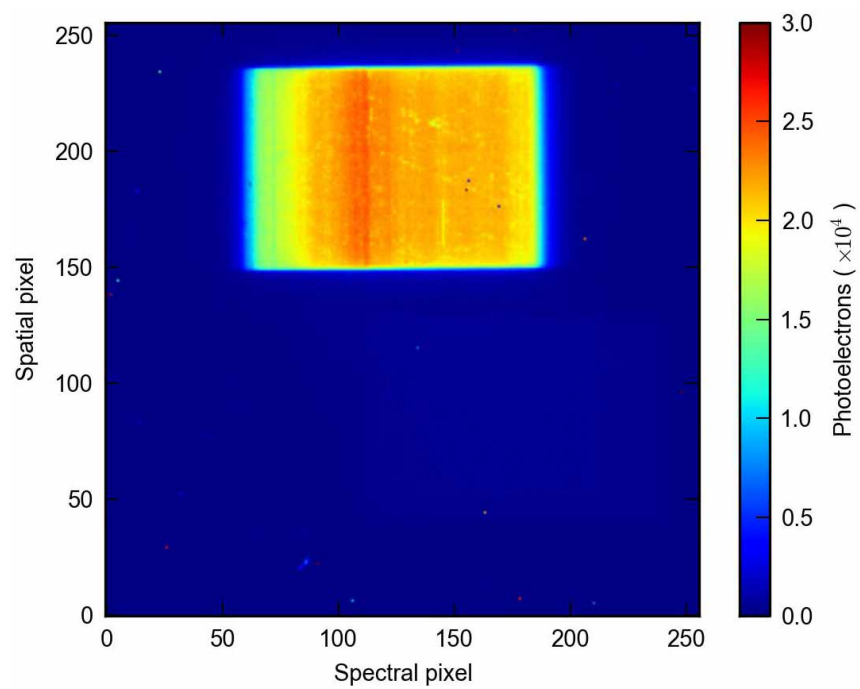
**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work is based on observations made with the NASA/ESA Hubble Space Telescope that were obtained at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract number NAS 5-26555. These observations are associated with program GO-13021. Support for this work was provided by NASA through a grant from the Space Telescope Science Institute, the National Science Foundation through a Graduate Research Fellowship (to L.K.), the Alfred P. Sloan Foundation through a Sloan Research Fellowship (to J.L.B.), NASA through a Sagan Fellowship (to J.-M.D.), and the European Research Council (for D.H. under the European Community's Seventh Framework Programme, FP7/2007-2013 Grant Agreement number 247060).

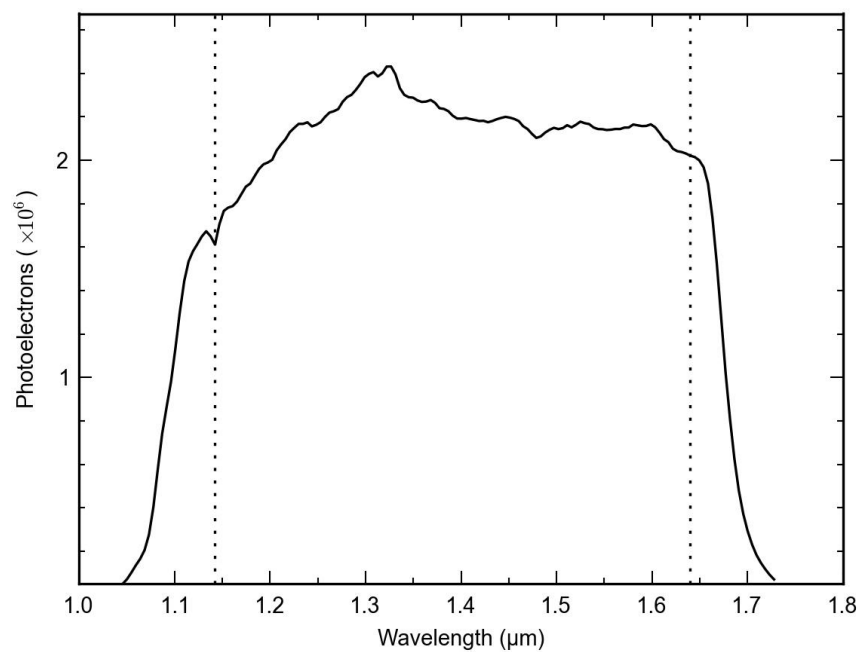
**Author Contributions** L.K. led the data analysis (with contributions from J.L.B., D.D., K.B.S. and A.S.). J.L.B. and J.-M.D. conceived the project and wrote the telescope time proposal (with contributions from B.B., D.D., S.S. and Z.B.-T.). L.K., J.L.B., J.-M.D., D.D. and Z.B.-T. planned the observations. B.B. and S.S. developed and performed the theoretical modelling. D.H. calculated the theoretical stellar limb darkening. J.L.B. led the overall direction of the project. L.K., J.L.B., J.-M.D. and B.B. wrote the paper. All authors discussed the results and commented on the manuscript.

**Author Information** The data used in this work can be accessed at the NASA Mikulski Archive for Space Telescopes (<http://archive.stsci.edu>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.K. ([laura.kreidberg@uchicago.edu](mailto:laura.kreidberg@uchicago.edu)).

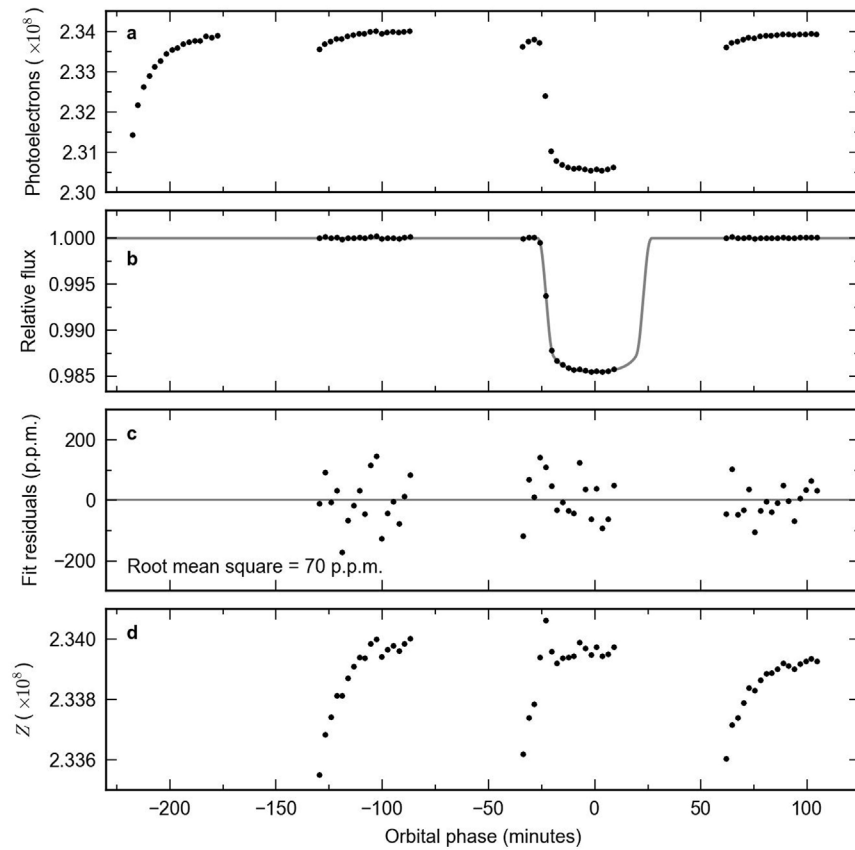




Extended Data Figure 1 | An example of a spatially scanned raw data frame. The exposure time was 88.4 s.

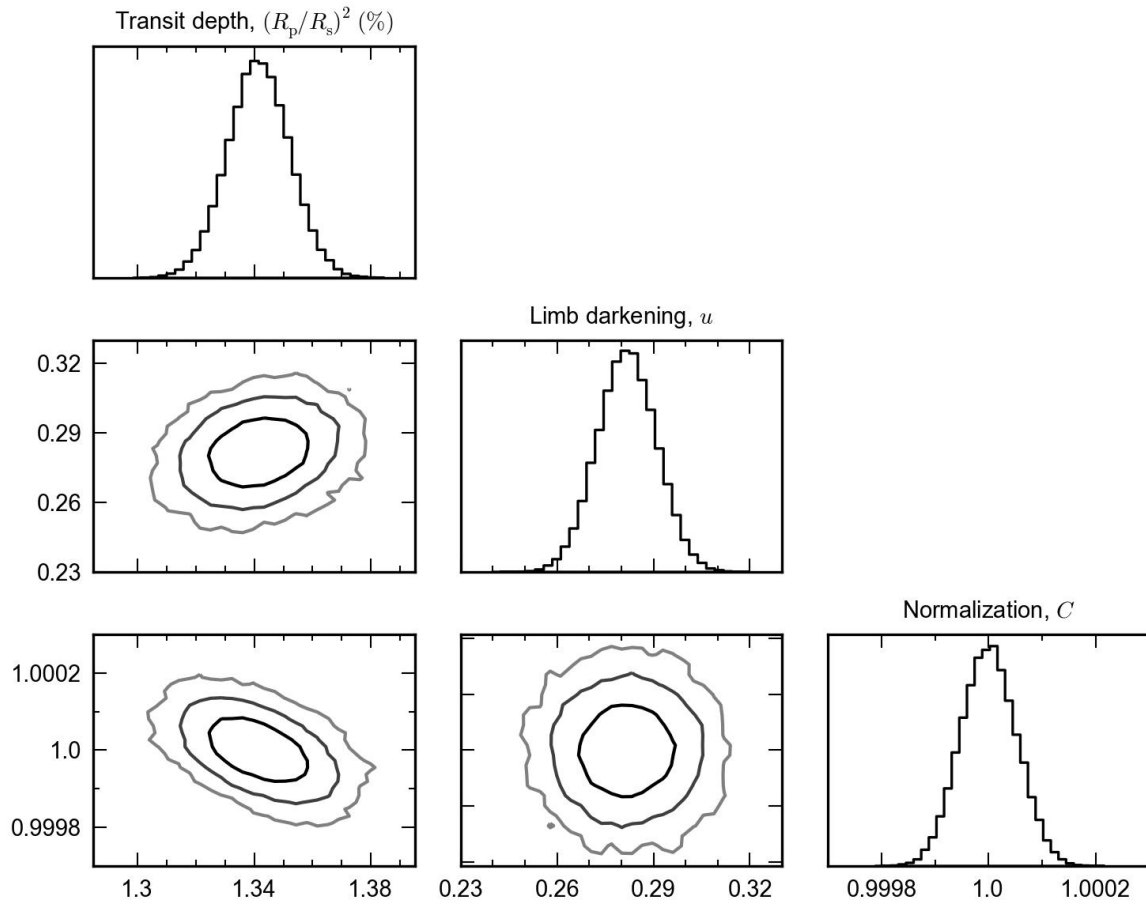


**Extended Data Figure 2** | An example of an extracted spectrum for an 88.4-s exposure. The dotted lines indicate the wavelength range over which we measure the transmission spectrum.



**Extended Data Figure 3 | The broadband light curve fit from the first transit observation.** **a**, The raw broadband light curve. **b**, The broadband light curve corrected for systematics using the model-ramp technique (points) and

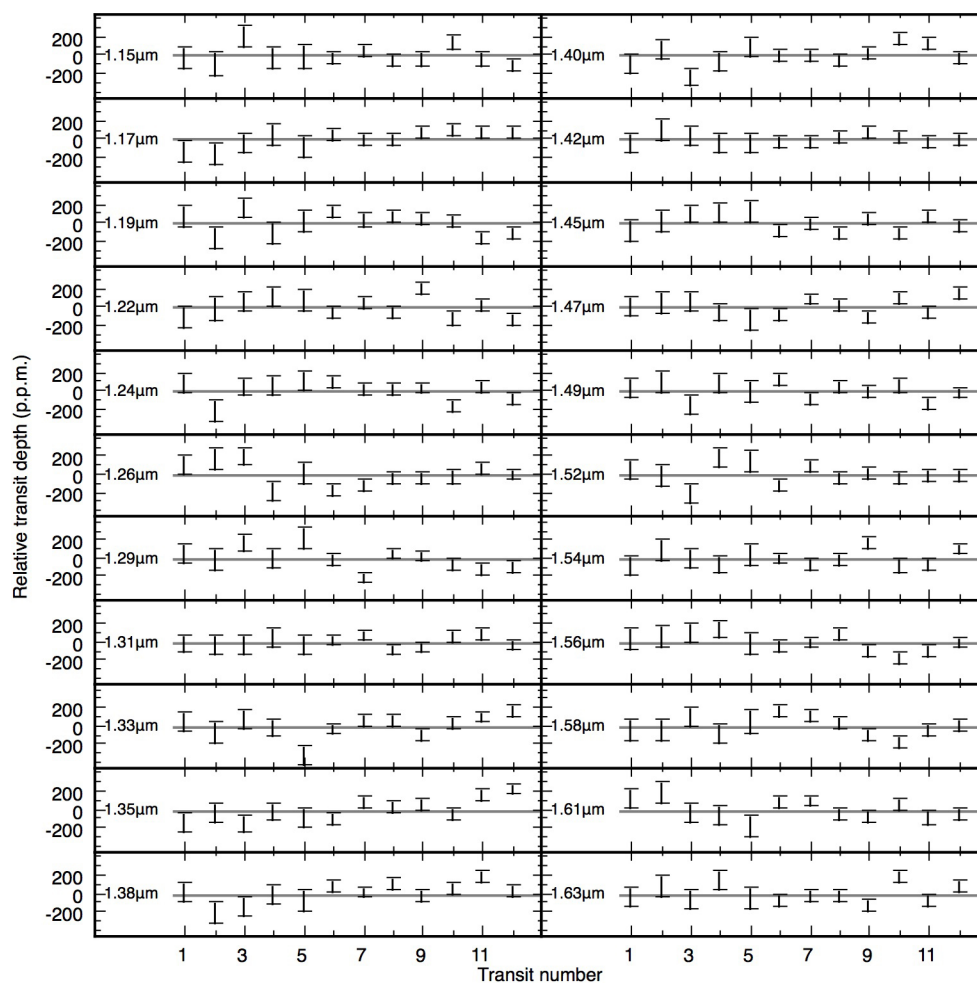
the best-fit model (line). **c**, Residuals from the broadband light curve fit. **d**, The vector of systematics  $Z$  (see the Supplementary Information) used in the divide-white technique.



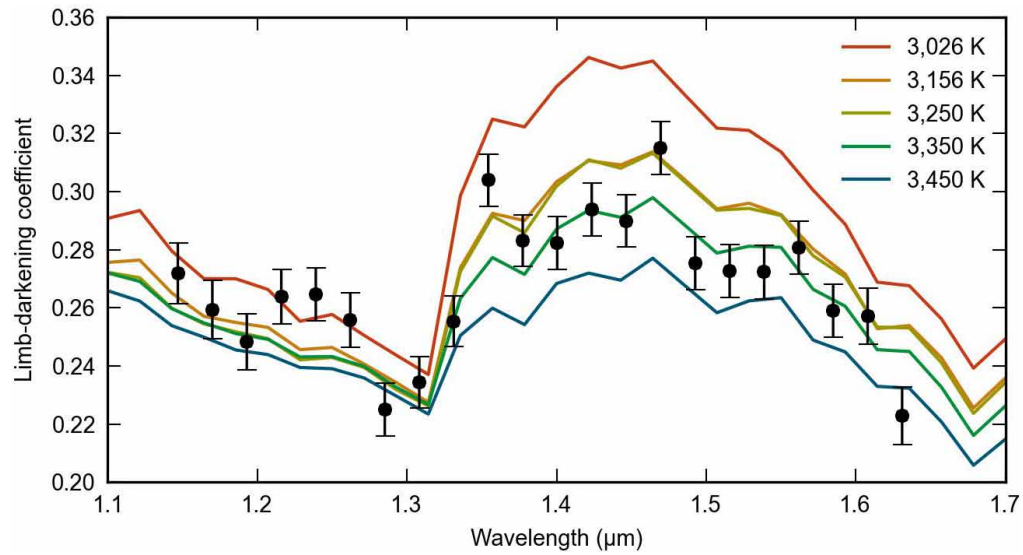
**Extended Data Figure 4 | The posterior distributions for the divide-white fit parameters for the 1.40- $\mu\text{m}$  channel from the first transit observation.** The histograms represent the Markov chains for each parameter. The contour plots

represent pairs of parameters, with lines indicating the  $1\sigma$ ,  $2\sigma$  and  $3\sigma$  confidence intervals for the distribution. The normalization constant is divided by its mean.





**Extended Data Figure 5 | Transit depths relative to the mean in 22 spectroscopic channels, for the 12 transits analysed.** The black error bars indicate the  $1\sigma$  uncertainties determined by a Markov chain Monte Carlo fit.



**Extended Data Figure 6 | Fitted limb-darkening coefficients as a function of wavelength (black points) and theoretical predictions for stellar atmospheres with a range of temperatures (lines). The uncertainties are  $1\sigma$**

confidence intervals from a Markov chain Monte Carlo fit. The temperature of GJ 1214 is estimated to be 3,250 K (ref. 22).

**Extended Data Table 1 | Derived parameters for the divide-white (d-w) and model-ramp (m-r) techniques**

Wavelength ( $\mu\text{m}$ )	Transit depth (p.p.m.)		Limb darkening		$\chi^2_\nu$	
	d-w	m-r	d-w	m-r	d-w	m-r
1.135 – 1.158	-39 $\pm$ 31	6 $\pm$ 33	0.27 $\pm$ 0.01	0.28 $\pm$ 0.01	1.12	1.20
1.158 – 1.181	-28 $\pm$ 30	12 $\pm$ 32	0.26 $\pm$ 0.01	0.27 $\pm$ 0.01	1.01	1.24
1.181 – 1.204	34 $\pm$ 30	29 $\pm$ 30	0.25 $\pm$ 0.01	0.26 $\pm$ 0.01	1.04	1.44
1.205 – 1.228	-48 $\pm$ 28	-32 $\pm$ 29	0.26 $\pm$ 0.01	0.28 $\pm$ 0.01	0.90	1.22
1.228 – 1.251	27 $\pm$ 28	25 $\pm$ 29	0.26 $\pm$ 0.01	0.28 $\pm$ 0.01	0.85	1.29
1.251 – 1.274	5 $\pm$ 27	-6 $\pm$ 29	0.26 $\pm$ 0.01	0.26 $\pm$ 0.01	0.97	1.29
1.274 – 1.297	13 $\pm$ 27	12 $\pm$ 27	0.23 $\pm$ 0.01	0.23 $\pm$ 0.01	1.00	1.50
1.297 – 1.320	14 $\pm$ 26	0 $\pm$ 27	0.23 $\pm$ 0.01	0.25 $\pm$ 0.01	0.96	1.38
1.320 – 1.343	29 $\pm$ 26	2 $\pm$ 28	0.26 $\pm$ 0.01	0.27 $\pm$ 0.01	1.08	1.52
1.343 – 1.366	-2 $\pm$ 27	-15 $\pm$ 28	0.30 $\pm$ 0.01	0.32 $\pm$ 0.01	0.99	1.44
1.366 – 1.389	32 $\pm$ 27	35 $\pm$ 26	0.28 $\pm$ 0.01	0.29 $\pm$ 0.01	0.97	1.42
1.389 – 1.412	31 $\pm$ 27	33 $\pm$ 28	0.28 $\pm$ 0.01	0.29 $\pm$ 0.01	0.96	1.39
1.412 – 1.435	-5 $\pm$ 27	-33 $\pm$ 28	0.29 $\pm$ 0.01	0.31 $\pm$ 0.01	1.15	1.51
1.435 – 1.458	29 $\pm$ 29	17 $\pm$ 28	0.29 $\pm$ 0.01	0.30 $\pm$ 0.01	1.01	1.39
1.458 – 1.481	-8 $\pm$ 28	1 $\pm$ 29	0.32 $\pm$ 0.01	0.33 $\pm$ 0.01	1.01	1.33
1.481 – 1.504	27 $\pm$ 28	28 $\pm$ 28	0.28 $\pm$ 0.01	0.29 $\pm$ 0.01	0.94	1.37
1.504 – 1.527	-11 $\pm$ 28	-23 $\pm$ 29	0.27 $\pm$ 0.01	0.29 $\pm$ 0.01	1.15	1.58
1.527 – 1.550	20 $\pm$ 28	1 $\pm$ 29	0.27 $\pm$ 0.01	0.29 $\pm$ 0.01	1.17	1.56
1.550 – 1.573	-21 $\pm$ 28	0 $\pm$ 28	0.28 $\pm$ 0.01	0.29 $\pm$ 0.01	1.20	1.62
1.573 – 1.596	-65 $\pm$ 28	-62 $\pm$ 30	0.26 $\pm$ 0.01	0.28 $\pm$ 0.01	1.08	1.46
1.596 – 1.619	-17 $\pm$ 28	-6 $\pm$ 29	0.26 $\pm$ 0.01	0.27 $\pm$ 0.01	1.34	1.69
1.619 – 1.642	-17 $\pm$ 30	-26 $\pm$ 30	0.22 $\pm$ 0.01	0.24 $\pm$ 0.01	1.16	1.59

# DNA-mediated nanoparticle crystallization into Wulff polyhedra

Evelyn Auyeung<sup>1,2</sup>, Ting I. N. G. Li<sup>1,2</sup>, Andrew J. Senesi<sup>2,3</sup>, Abrin L. Schmucker<sup>2,3</sup>, Bridget C. Pals<sup>2</sup>, Monica Olvera de la Cruz<sup>1,2,3</sup> & Chad A. Mirkin<sup>1,2,3</sup>

Crystallization is a fundamental and ubiquitous process much studied over the centuries. But although the crystallization of atoms is fairly well understood<sup>1,2</sup>, it remains challenging to predict reliably the outcome of molecular crystallization processes that are complicated by various molecular interactions and solvent involvement. This difficulty also applies to nanoparticles: high-quality three-dimensional crystals<sup>3–6</sup> are mostly produced using drying and sedimentation techniques that are often impossible to rationalize and control to give a desired crystal symmetry, lattice spacing and habit (crystal shape). In principle, DNA-mediated assembly of nanoparticles offers an ideal opportunity for studying nanoparticle crystallization<sup>7–17</sup>: a well-defined set of rules have been developed to target desired lattice symmetries and lattice constants<sup>8,9,18</sup>, and the occurrence of features such as grain boundaries and twinning in DNA superlattices and traditional crystals comprised of molecular or atomic building blocks suggests that similar principles govern their crystallization. But the presence of charged biomolecules, interparticle spacings of tens of nanometres, and the realization so far of only polycrystalline DNA-interconnected nanoparticle superlattices, all suggest that DNA-guided crystallization may differ from traditional crystal growth. Here we show that very slow cooling, over several days, of solutions of complementary-DNA-modified nanoparticles through the melting temperature of the system gives the thermodynamic product with a specific and uniform crystal habit. We find that our nanoparticle assemblies have the Wulff equilibrium crystal structure that is predicted from theoretical considerations and molecular dynamics simulations, thus establishing that DNA hybridization can direct nanoparticle assembly along a pathway that mimics atomic crystallization.

The crystallization of nanoparticles mediated by DNA typically involves initial assembly of a disordered aggregate, which upon thermal annealing slightly below its melting temperature transforms into an ordered superlattice (Fig. 1a, blue arrows)<sup>12</sup>. Transmission electron microscopy (TEM) images and the presence of rings in the small-angle X-ray scattering (SAXS) data show that all superlattices formed using this approach thus far are polycrystalline, with ordered micrometre-sized domains randomly oriented with respect to one another<sup>9,19</sup>. Considering that traditional crystallization techniques for atoms and molecules typically rely on slow cooling through the melting temperature<sup>20</sup>, we hypothesized that such a slow cooling approach applied to DNA-based assembly strategies might yield faceted crystals. DNA-functionalized nanoparticle solutions were therefore heated to above the melting temperature of the DNA links designed to connect particles and then slowly cooled to room temperature (Fig. 1a, red arrows), in a process which typically took two to three days to complete. A key parameter known to the researcher before doing the experiments is the aggregate melting temperature, which is well defined and directly correlated with the nucleic acid sequences used for assembly<sup>12</sup>.

The slow cooling of the combination of two sets of gold nanoparticles functionalized with complementary DNA linker strands<sup>9</sup> produces

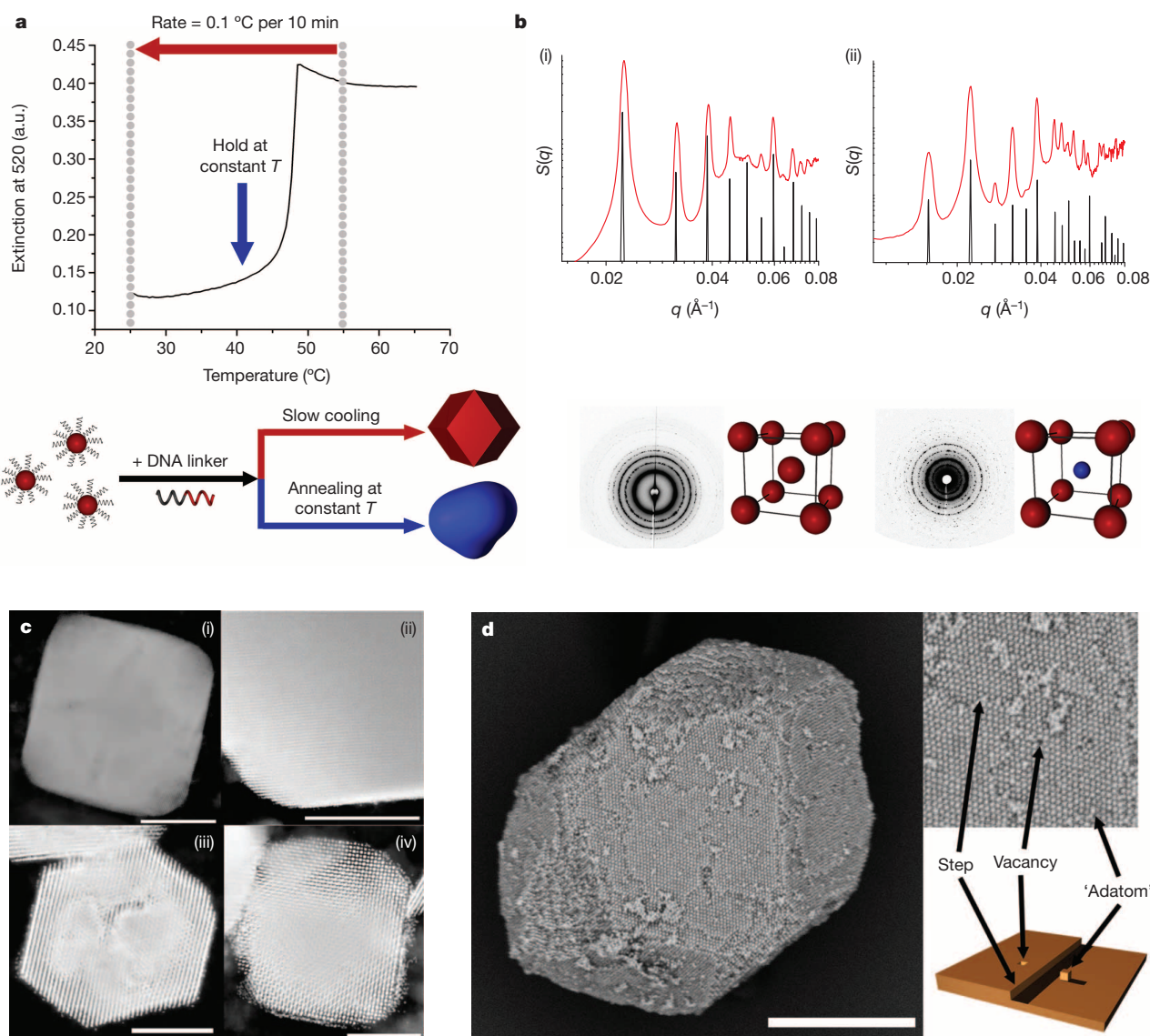
superlattices with the expected body-centred cubic (b.c.c.) packing when using 20-nm gold nanoparticles and CsCl packing when using 20-nm and 15-nm gold nanoparticles, as confirmed by the radially averaged one-dimensional SAXS data (Fig. 1b, (i) and (ii)). To enable direct visualization, the superlattices were also stabilized by embedding in silica<sup>19</sup>. TEM images of these structures reveal uniform crystals with square- and hexagonal-shaped domains for both the b.c.c. (Fig. 1c, (i) and (ii)) and the CsCl (Fig. 1c, (iii) and (iv)) particle packing symmetries, whereas scanning electron microscopy (SEM) allows us to observe surface features and the overall crystal habit (Fig. 1d). Evidently, the slow-cooling process enables DNA-driven assembly and crystallization that favours faceted rhombic dodecahedron microcrystals over the polycrystalline assemblies obtained by annealing below the melting temperature. Although single-crystal formation by annealing below the melting temperature may in principle be possible, the kinetics of reorganization from an irregularly shaped crystal into a well-defined microcrystal are likely to be too slow to be observed experimentally.

SEM images of the microcrystals in different orientations on the substrate are all consistent with rhombic dodecahedron formation (Fig. 2a). Closer inspection of one of the crystals (Fig. 2a, bottom) reveals extraordinarily well ordered nanoparticles at the surface, as well as the presence of common surface defects including ‘particle adatoms’ (a surface defect in which an atom is adsorbed on the surface of a crystal plane) and step edges. The nanoparticle orientation is consistent with a crystal that is enclosed by (110) planes, as expected for rhombic dodecahedra, and which is also the closest-packed plane in a b.c.c. unit cell. A tilting experiment was conducted in the TEM on a single microcrystal to observe the different morphologies that are consistent with a rhombic dodecahedron crystal habit. It is important to note that although these microcrystals exhibited a wide size distribution (Fig. 2c), faceted crystals were the predominant product of slow cooling, and no shapes other than rhombic dodecahedron microcrystals were observed.

In contrast to several prior examples of microcrystals grown from nanoparticle building blocks, in which the overall crystal shape was largely dependent on factors including nanoparticle size and length of the ligand shell<sup>3,4,21</sup>, the shape of the microcrystals we report here was fairly independent of such parameters: microcrystals made from 5-nm, 10-nm and 20-nm gold nanoparticles all exhibit overall rhombic dodecahedron shapes and b.c.c. packing with lattice parameters of 25.7 nm, 29.1 nm and 39.5 nm, respectively (compare Fig. 3a, b and d). Oligonucleotide length was kept constant in these experiments. Furthermore, rhombic dodecahedron microcrystals were observed for a binary system consisting of 20-nm and 15-nm gold nanoparticles arranged in a CsCl lattice symmetry (Fig. 3c). Thus, we conclude that the rhombic dodecahedron is the thermodynamically most favourable crystal shape for this system over a range of particle sizes and interparticle distances. Furthermore, molecular dynamics simulations on a colloid model predicted a rhombic dodecahedron equilibrium crystal structure, fully consistent with experimental observations (Fig. 3e).

<sup>1</sup>Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, USA. <sup>2</sup>International Institute for Nanotechnology, Northwestern University, Evanston, Illinois 60208, USA. <sup>3</sup>Department of Chemistry, Northwestern University, Evanston, Illinois 60208, USA.





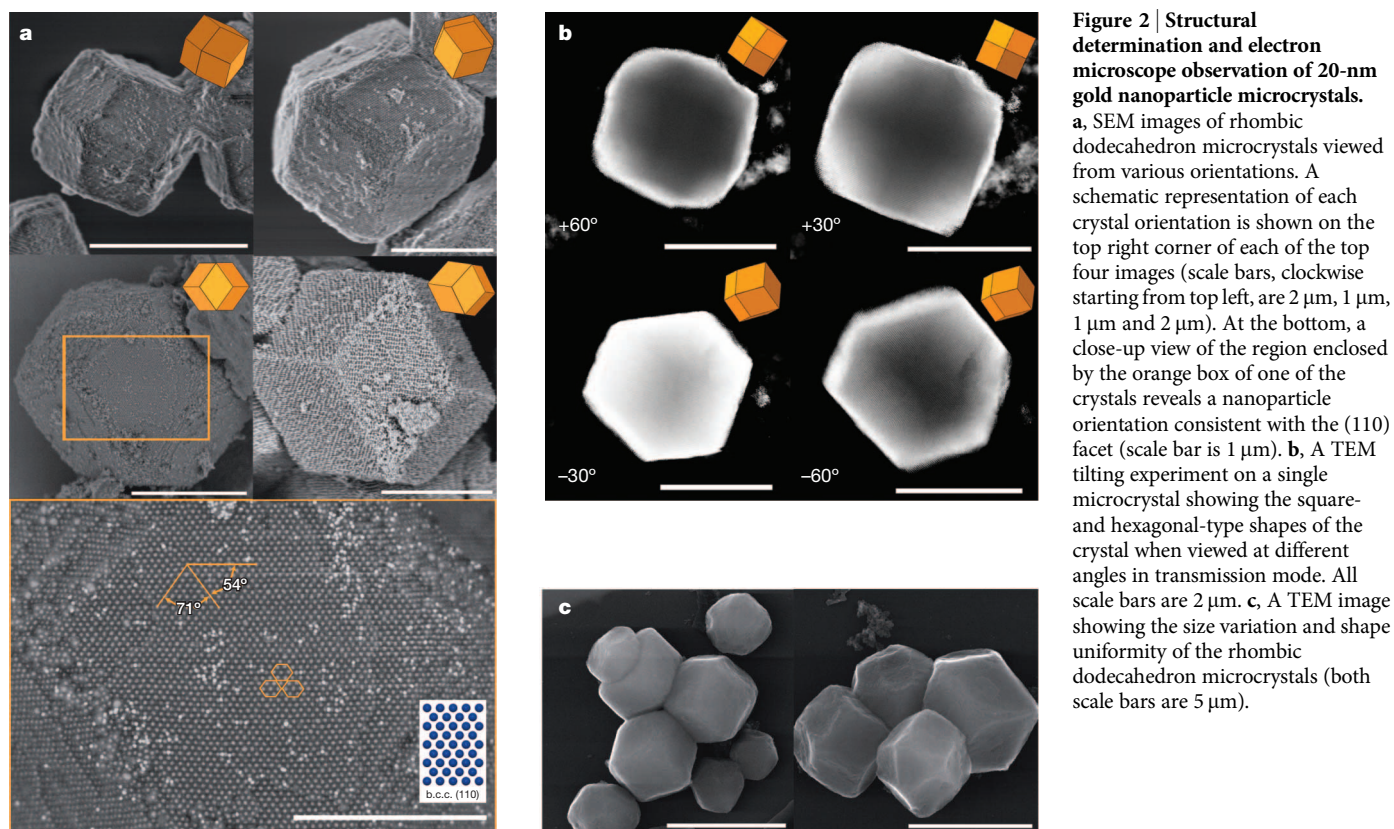
**Figure 1 | Superlattices formed by the slow-cooling method.** **a**, Two approaches for DNA-mediated nanoparticle crystallization are designated by arrows on a thermal melting curve of a DNA-linked gold nanoparticle aggregate. The traditional method (blue arrow), in which the aggregate is annealed a few degrees below the melting temperature, produces polycrystalline superlattices with no defined shape. The slow-cooling method (red arrow), in which the aggregate is heated above its melting temperature and cooled at a rate of  $0.01\text{ }^{\circ}\text{C min}^{-1}$ , produced well-defined, faceted microcrystals in each of the dozens of experiments conducted using these conditions. Extinction wavelength measured from the ultraviolet–visible spectrum, 520 nm (a.u.). **b**, Representative one-dimensional (top) and two-dimensional (bottom)

SAXS data for b.c.c. (i) and CsCl (ii) superlattices synthesized from the slow-cooling technique. In the one-dimensional data, the red trace is the experimentally obtained scattering pattern and the black trace is the predicted scattering pattern for a perfect lattice. SAXS data are shown as plots of superlattice structure factor ( $S(q)$ , in arbitrary units) versus scattering vector ( $q$ , in units of  $\text{\AA}^{-1}$ ). **c**, TEM images of shape-controlled b.c.c. (i)–(ii) and CsCl (iii)–(iv) microcrystals. Scale bars are  $1.5\text{ }\mu\text{m}$  for (i),  $1.0\text{ }\mu\text{m}$  for (ii),  $0.5\text{ }\mu\text{m}$  for (iii) and  $0.5\text{ }\mu\text{m}$  for (iv). **d**, SEM image of a representative b.c.c. microcrystal with visible faceting where constituent nanoparticles can readily be seen (20-nm gold nanoparticles shown; scale bar is  $1\text{ }\mu\text{m}$ ). The inset shows a high-magnification ( $\times 52,500$ ) view of the crystal facet with labelled surface defects.

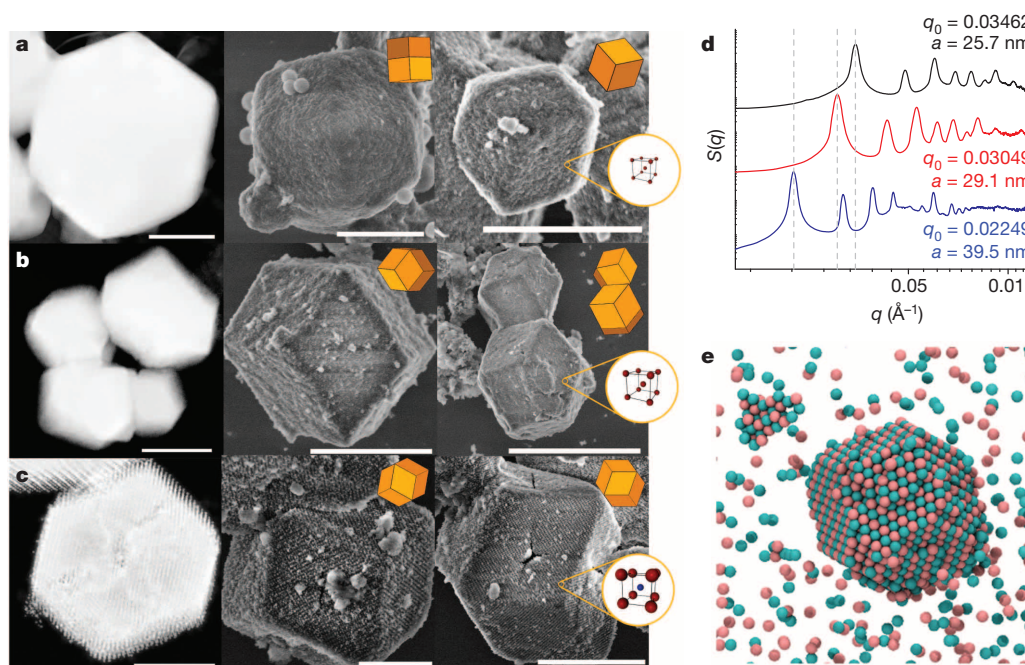
The formation of rhombic dodecahedron microcrystals from a b.c.c. packing of nanoparticles can be rationalized in terms of the surface energy  $\gamma$  of the exposed facets<sup>22</sup>. Rhombic dodecahedra are enclosed by (110) facets (bottom panel of Fig. 2a), which is the closest-packed plane for a b.c.c. or CsCl lattice. When using the standard broken-bond model approximation for surface energy, exposing the closest-packed plane is thermodynamically favoured: it requires breaking the smallest number of particle-to-particle interactions per unit area and thus exposes the lowest-surface-energy facet<sup>23</sup>. From this model, the relative surface energies for b.c.c. metal facets should exhibit a ratio of  $\gamma_{(110)}:\gamma_{(111)}:\gamma_{(100)} = 1:1.22:1.41$ . Similarly, the relative surface energies for face-centred cubic (f.c.c.) metal facets should be

$\gamma_{(111)}:\gamma_{(100)}:\gamma_{(110)} = 1:1.15:1.22$ . These calculations thus predict the Wulff polyhedron, the equilibrium crystal structure, to be a rhombic dodecahedron enclosed by (110) facets for a b.c.c. metal and a truncated octahedron enclosed by (111) and (100) facets for a f.c.c. metal<sup>1</sup>.

In many systems the expected Wulff polyhedron is not always formed, and the validity of the assumptions and approximations made must be analysed for each individual case. We therefore calculated actual surface energy values for our DNA–nanoparticle system using recently developed molecular dynamics simulations that accurately predict the crystallization behaviour of DNA-assembled nanoparticles (see Supplementary Information)<sup>24,25</sup>. In these calculations, the surface energy is defined as the excess energy at the surface of a material



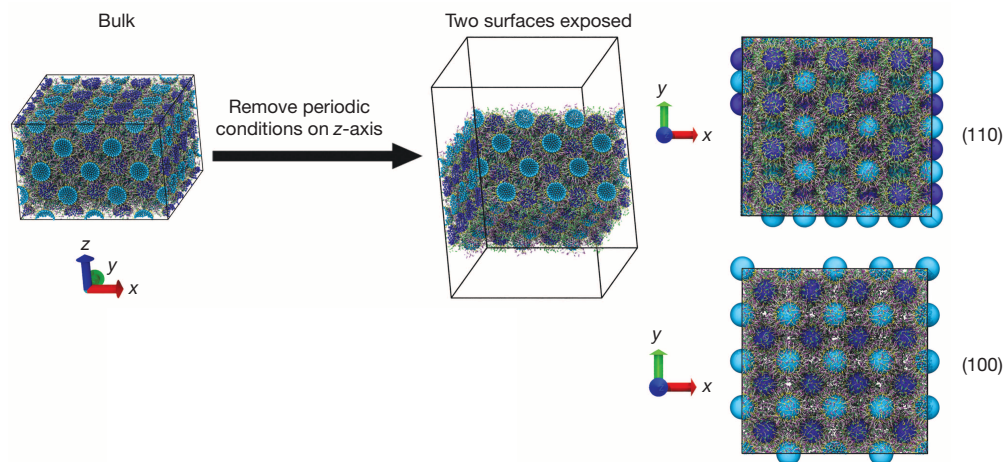
**Figure 2 | Structural determination and electron microscope observation of 20-nm gold nanoparticle microcrystals.** **a**, SEM images of rhombic dodecahedron microcrystals viewed from various orientations. A schematic representation of each crystal orientation is shown on the top right corner of each of the top four images (scale bars, clockwise starting from top left, are 2  $\mu\text{m}$ , 1  $\mu\text{m}$ , 1  $\mu\text{m}$  and 2  $\mu\text{m}$ ). At the bottom, a close-up view of the region enclosed by the orange box of one of the crystals reveals a nanoparticle orientation consistent with the (110) facet (scale bar is 1  $\mu\text{m}$ ). **b**, A TEM tilting experiment on a single microcrystal showing the square- and hexagonal-type shapes of the crystal when viewed at different angles in transmission mode. All scale bars are 2  $\mu\text{m}$ . **c**, A TEM image showing the size variation and shape uniformity of the rhombic dodecahedron microcrystals (both scale bars are 5  $\mu\text{m}$ ).



**Figure 3 | Rhombic dodecahedron microcrystals with varying unit cell compositions.** **a–c**, TEM and SEM images of rhombic dodecahedron microcrystals synthesized from a b.c.c. lattice of 5-nm gold nanoparticles (scale bars, left to right, are 1  $\mu\text{m}$ , 1  $\mu\text{m}$  and 2  $\mu\text{m}$ ) (**a**), from a b.c.c. lattice of 10-nm gold nanoparticles (scale bars, left to right, are 1  $\mu\text{m}$ , 2  $\mu\text{m}$  and 4  $\mu\text{m}$ ) (**b**) and from a CsCl lattice of 20-nm and 15-nm gold nanoparticles (scale bars, left to right, are 0.5  $\mu\text{m}$ , 1  $\mu\text{m}$  and 2  $\mu\text{m}$ ) (**c**). **d**, SAXS data for a b.c.c. crystal made

from 5-nm (black trace), 10-nm (red trace), and 20-nm (blue trace) gold nanoparticles. First-order scattering peak  $q_0$  and corresponding lattice parameter values are indicated next to the respective scattering pattern for each crystal. **e**, Molecular dynamics simulation of a binary set of particles exhibiting interactions modelled for the DNA–gold nanoparticle system produces a rhombic dodecahedron microcrystal that is consistent with experimental observations.





**Figure 4 | Method for calculating surface energy values for a b.c.c. DNA-gold nanoparticle superlattice.** Periodic boundary conditions along the z-axis are removed from the bulk crystal to expose the surface of interest, for example,

compared to the energy of the bulk system. To calculate  $\gamma$ , periodic boundary conditions were removed from the modelled bulk crystal along the z-axis to expose the facet of interest on two sides (Fig. 4). The energy of the bulk crystal  $E_{\text{bulk}}$  was subtracted from the energy of the exposed facet  $E_{\text{surface exposed}}$ , and then divided by twice the area (because two surfaces were exposed and the surface charge density is close to zero<sup>26</sup>) to give the surface energy  $\gamma$  of the exposed facet.

Table 1 summarizes absolute surface-energy values for the facets of a b.c.c. crystal and a f.c.c. crystal consisting of DNA-assembled nanoparticles calculated using this model, with the binding strength of complementary sticky ends scaled to  $42.3 \text{ kJ mol}^{-1}$  (ref. 27) to match the strength of the DNA sequences (TTCCTT) used in our experiments. For the b.c.c. system, the calculated ratios of  $\gamma_{(100)}:\gamma_{(110)} = 1.46 \pm 0.02$  and  $\gamma_{(111)}:\gamma_{(110)} = 1.24 \pm 0.02$  are in good agreement with the theoretically predicted ratio described above. Evidently, the observation of uniform rhombic dodecahedron crystals from a b.c.c. arrangement of nanoparticles follows the crystallization behaviour expected for a b.c.c. arrangement of atoms<sup>28</sup>. The expected Wulff polyhedron was observed for the b.c.c. nanoparticle system, but no truncated octahedra or other uniform shapes were observed in either experiment or simulation among the faceted crystals obtained for the f.c.c. system. This is probably because the surface energies of the two most stable surfaces in a f.c.c. crystal are too close in energy (predicted and calculated ratio  $\gamma_{(100)}:\gamma_{(111)} = 1.15$ ) for one to be favoured predominantly over the other (Table 1). Furthermore, the SAXS data for the f.c.c. crystals showed evidence of stacking faults and twinning in the lattice structure, defects which may have prevented the formation of uniform crystal shapes (a more in-depth discussion of the f.c.c. system can be found in the Supplementary Information). Nonetheless, the consistency between the experimental observations and the simulation results provides convincing evidence that the broken-bond approximation used for describing surface energy

(110) or (100) as shown. Absolute surface energy values calculated from this approach are found in Table 1.  $\gamma = (E_{\text{surface exposed}} - E_{\text{bulk}})/(2 \times \text{area})$ .

and crystal growth for atomic systems can similarly be used to describe the crystallization of nanoparticles using DNA interactions; and, hence, that the DNA-guided assembly of nanoparticles provides a nanometre-scale analogue to the crystallization behaviour exhibited by atomic crystals.

The experimental observation of the Wulff equilibrium crystal structure, coupled with computational models, demonstrates the utility of DNA for controlling not only the recognition properties and surface energy of individual nanoparticles, but also the surface energies of the macroscopic nanoparticle assembly in such way that a specific structure can be deliberately programmed and realized in the laboratory. The challenge now for both the experimental and theoretical communities is to build on the principles we have described here to identify and synthesize crystal habits that maximize surface-energy differences, and to create single microcrystals with useful properties that may find practical use such as in photonic and catalytic applications.

## METHODS SUMMARY

All oligonucleotides used in this work were synthesized on a solid-support MM48 synthesizer using reagents purchased from Glen Research. Sequences can be found in the Supplementary Information. Nanoparticles were functionalized and assembled according to published literature protocols. After particle assembly, slow cooling to room temperature was conducted in a temperature cycler (Life Technologies) at a starting temperature above the aggregate melting temperature and typically at a rate of  $0.01^\circ \text{C min}^{-1}$  unless otherwise specified. Superlattices were characterized by synchrotron SAXS experiments conducted at the Advanced Photon Source at Argonne National Laboratory. Superlattices were transferred to the solid state using a silica embedding method<sup>19</sup> for visualization by TEM (Hitachi HD2300) and SEM (Hitachi SU8030). To reproduce the shapes with molecular dynamics simulations, a colloidal model was validated by computing the interaction potential with simulations<sup>24,26</sup> with explicit DNA chains and simulations were performed with the LAMMPS package (available at <http://lammps.sandia.gov/>). See the

**Table 1 | Surface energy values calculated for DNA-gold nanoparticle superlattices**

Structure	System	Relaxed first-neighbour distance (nm)	Lattice constant (nm)	$\gamma_{(100)}$ (mJ m <sup>-2</sup> )	$\gamma_{(110)}$ (mJ m <sup>-2</sup> )	$\gamma_{(111)}$ (mJ m <sup>-2</sup> )	Ratios of surface energies
Body-centred cubic	20-nm gold nanoparticle, 150 strands, 18 bp	34.0	39.3 (simulation) 39.5 (experiment)	$0.548 \pm 0.005$	$0.375 \pm 0.005$	$0.464 \pm 0.003$	(110):(100):(111) = 1: 1.46 $\pm$ 0.02: 1.24 $\pm$ 0.02
Face-centred cubic	20-nm gold nanoparticle, 150 strands, 43 bp	47.8	67.6 (simulation) 67.4 (experiment)	$0.094 \pm 0.006$	$0.104 \pm 0.002$	$0.082 \pm 0.005$	(111):(100):(110) = 1: 1.15 $\pm$ 0.10: 1.27 $\pm$ 0.08

Surface-energy values consider only the contribution of DNA hybridization. As a reference, the surface energy of water is  $72 \text{ mJ m}^{-2}$ . bp, number of DNA base pairs. The stated values and errors indicate the average and associated standard deviation of approximately five independent simulation runs with random initial conditions. The relaxed lattice parameter from simulation is shown as the topmost value; the lattice parameter from experiment is shown as the lowermost value.

Supplementary Videos for a visualization of the simulated formation of micro-crystals from a b.c.c. and a f.c.c. system of interacting particles modelled as a single bead. To estimate surface-energy values, a scale-accurate coarse-grained model was used<sup>24</sup> and molecular dynamics simulations were performed on the HOOMD-Blue package (available at <http://codeblue.umich.edu/hoomd-blue>). More molecular dynamics simulation details and assumptions can be found in the Supplementary Information.

Received 9 July; accepted 27 September 2013.

Published online 27 November 2013.

- Wulff, G. On the question of speed of growth and dissolution of crystal surfaces. *Z. Kristallogr. Mineral.* **34**, 449–530 (1901).
- Xia, Y. N., Xiong, Y. J., Lim, B. & Skrabalak, S. E. Shape-controlled synthesis of metal nanocrystals: simple chemistry meets complex physics? *Angew. Chem. Int. Edn* **48**, 60–103 (2009).
- Rupich, S. M., Shevchenko, E. V., Bodnarchuk, M. I., Lee, B. & Talapin, D. V. Size-dependent multiple twinning in nanocrystal superlattices. *J. Am. Chem. Soc.* **132**, 289–296 (2010).
- Compton, O. C. & Osterloh, F. E. Evolution of size and shape in the colloidal crystallization of gold nanoparticles. *J. Am. Chem. Soc.* **129**, 7793–7798 (2007).
- Kalsin, A. M. *et al.* Electrostatic self-assembly of binary nanoparticle crystals with a diamond-like lattice. *Science* **312**, 420–424 (2006).
- Wang, T., LaMontagne, D., Lynch, J., Zhuang, J. Q. & Cao, Y. C. Colloidal superparticles from nanoparticle assembly. *Chem. Soc. Rev.* **42**, 2804–2823 (2013).
- Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607–609 (1996).
- Macfarlane, R. J., O'Brien, M. N., Hurst, S. J. & Mirkin, C. A. Nucleic acid-modified nanostructures as programmable atom equivalents: forging a new “table of elements”. *Angew. Chem. Int. Edn* **52**, 5688–5698 (2013).
- Macfarlane, R. J. *et al.* Nanoparticle superlattice engineering with DNA. *Science* **334**, 204–208 (2011).
- Auyeung, E. *et al.* Synthetically programmable nanoparticle superlattices using a hollow three-dimensional spacer approach. *Nature Nanotechnol.* **7**, 24–28 (2012).
- Jones, M. R. *et al.* DNA-nanoparticle superlattices formed from anisotropic building blocks. *Nature Mater.* **9**, 913–917 (2010).
- Park, S. Y. *et al.* DNA-programmable nanoparticle crystallization. *Nature* **451**, 553–556 (2008).
- Nykypanchuk, D., Maye, M. M., van der Lelie, D. & Gang, O. DNA-guided crystallization of colloidal nanoparticles. *Nature* **451**, 549–552 (2008).
- Cutler, J. I., Auyeung, E. & Mirkin, C. A. Spherical nucleic acids. *J. Am. Chem. Soc.* **134**, 1376–1391 (2012).
- Zhang, C. *et al.* A general approach to DNA-programmable atom equivalents. *Nature Mater.* **12**, 741–746 (2013).
- Senesi, A. J. *et al.* Stepwise evolution of DNA-programmable nanoparticle superlattices. *Angew. Chem. Int. Edn* **52**, 6624–6628 (2013).
- Macfarlane, R. J., Jones, M. R., Lee, B., Auyeung, E. & Mirkin, C. A. Topotactic interconversion of nanoparticle superlattices. *Science* **341**, 1222–1225 (2013).
- Macfarlane, R. J. *et al.* Establishing the design rules for DNA-mediated colloidal crystallization. *Angew. Chem. Int. Edn* **49**, 4589–4592 (2010).
- Auyeung, E., Macfarlane, R. J., Choi, C. H. J., Cutler, J. I. & Mirkin, C. A. Transitioning DNA-engineered nanoparticle superlattices from solution to the solid state. *Adv. Mater.* **24**, 5181–5186 (2012).
- Abrosimov, N. V., Rossolenko, S. N., Thieme, W., Gerhardt, A. & Schroder, W. Czochralski growth of Si- and Ge-rich SiGe single crystals. *J. Cryst. Growth* **174**, 182–186 (1997).
- Nagaoka, Y., Chen, O., Wang, Z. W. & Cao, Y. C. Structural control of nanocrystal superlattices using organic guest molecules. *J. Am. Chem. Soc.* **134**, 2868–2871 (2012).
- Mackenzie, J. K., Moore, A. J. W. & Nicholas, J. F. Bonds broken at atomically flat crystal surfaces. 1. Face-centred and body-centred cubic crystals. *J. Phys. Chem. Solids* **23**, 185–196 (1962).
- Sundquist, B. E. Direct determination of anisotropy of surface free energy of solid gold, silver, copper, nickel, and alpha and gamma iron. *Acta Metall. Mater.* **12**, 67–86 (1964).
- Li, T. I. N. G., Sknepnek, R., Macfarlane, R. J., Mirkin, C. A. & de la Cruz, M. O. Modeling the crystallization of spherical nucleic acid nanoparticle conjugates with molecular dynamics simulations. *Nano Lett.* **12**, 2509–2514 (2012).
- Knorowski, C., Burleigh, S. & Travesset, A. Dynamics and statics of DNA-programmable nanoparticle self-assembly and crystallization. *Phys. Rev. Lett.* **106**, 215501–215504 (2011).
- Zwanikken, J. W., Guo, P. J., Mirkin, C. A. & de la Cruz, M. O. Local ionic environment around polyvalent nucleic acid-functionalized nanoparticles. *J. Phys. Chem. C* **115**, 16368–16373 (2011).
- Li, T. I. N. G., Sknepnek, R. & de la Cruz, M. O. Thermally active hybridization drives the crystallization of DNA-functionalized nanoparticles. *J. Am. Chem. Soc.* **135**, 8535–8541 (2013).
- Che, J. G., Chan, C. T., Jian, W. E. & Leung, T. C. Surface atomic structures, surface energies, and equilibrium crystal shape of molybdenum. *Phys. Rev. B* **57**, 1875–1880 (1998).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** C.A.M. and M.O.d.I.C. acknowledge support from the following awards: the Air Force Office of Scientific Research (AFOSR) Multidisciplinary University Research Initiative (MURI) FA9550-11-1-0275, the National Science Foundation Materials Research Science and Engineering Center programme DMR-1121262 at the Materials Research Center of Northwestern University, and the Non-equilibrium Energy Research Center (NERC), an Energy Frontier Research Center funded by the Department of Energy (DoE), Office of Science, Office of Basic Energy Sciences under Award DE-SC0000989. E.A. acknowledges a National Defense Science and Engineering Graduate (NDSEG) Fellowship (number 32 CFR 168a). T.L. acknowledges a Ryan Fellowship from Northwestern University. T.L. thanks S. Dhakal and K. Kohlstedt for sharing scripts on the colloid model and S. Patala and J. Zwanikken for discussions. SAXS experiments were carried out at the Dupont–Northwestern–Dow Collaborative Access Team beam line at the Advanced Photon Source (APS) at Argonne National Laboratory, and use of the APS was supported by the DoE (DE-AC02-06CH11357). The electron microscopy work was performed at the Electron Probe Instrumentation Center of the NU Atomic and Nanoscale Characterization Experimental Center at Northwestern University. The computational work was performed using the TARDIS computer cluster supported by the US Department of Defense National Security Science and Engineering Faculty Fellowship (number FA9550-10-1-0167).

**Author Contributions** E.A., T.I.N.G.L., A.J.S., M.O.d.I.C. and C.A.M. designed experiments and analysed data. E.A. collected and analysed data for electron microscopy and X-ray studies. T.I.N.G.L. collected molecular dynamics simulation results. T.I.N.G.L. and M.O.d.I.C. wrote the theoretical model and the simulation details found in the Supplementary Information. E.A. and C.A.M. wrote the manuscript. A.J.S., A.L.S. and B.P. prepared samples and collected electron microscopy data.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.A.M. ([chadnano@northwestern.edu](mailto:chadnano@northwestern.edu)).



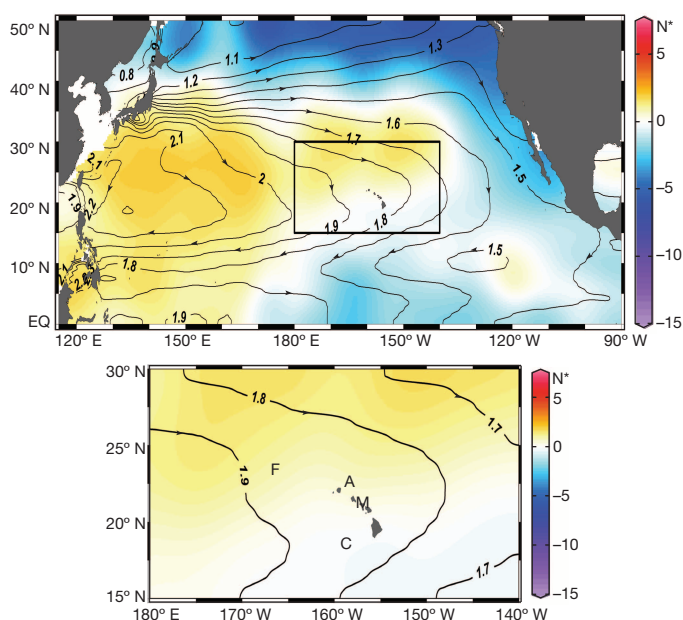
# Increasing subtropical North Pacific Ocean nitrogen fixation since the Little Ice Age

Owen A. Sherwood<sup>1†</sup>, Thomas P. Guilderson<sup>1,2,3</sup>, Fabian C. Batista<sup>1</sup>, John T. Schiff<sup>1</sup> & Matthew D. McCarthy<sup>1</sup>

The North Pacific subtropical gyre (NPSG) plays a major part in the export of carbon and other nutrients to the deep ocean<sup>1</sup>. Primary production in the NPSG has increased in recent decades despite a reduction in nutrient supply to surface waters<sup>2,3</sup>. It is thought that this apparent paradox can be explained by a shift in plankton community structure from mostly eukaryotes to mostly nitrogen-fixing prokaryotes<sup>2–4</sup>. It remains uncertain, however, whether the plankton community domain shift can be linked to cyclical climate variability or a long-term global warming trend<sup>5</sup>. Here we analyse records of bulk and amino-acid-specific  $^{15}\text{N}/^{14}\text{N}$  isotopic ratios ( $\delta^{15}\text{N}$ ) preserved in the skeletons of long-lived deep-sea proteinaceous corals collected from the Hawaiian archipelago; these isotopic records serve as a proxy for the source of nitrogen-supported export production through time. We find that the recent increase in nitrogen fixation is the continuation of a much larger, centennial-scale trend. After a millennium of relatively minor fluctuation,  $\delta^{15}\text{N}$  decreases between 1850 and the present. The total shift in  $\delta^{15}\text{N}$  of  $-2$  per mil over this period is comparable to the total change in global mean sedimentary  $\delta^{15}\text{N}$  across the Pleistocene–Holocene transition, but it is happening an order of magnitude faster<sup>6</sup>. We use a steady-state model and find that the isotopic mass balance between nitrate and nitrogen fixation implies a 17 to 27 per cent increase in nitrogen fixation over this time period. A comparison with independent records<sup>7,8</sup> suggests that the increase in nitrogen fixation might be linked to Northern Hemisphere climate change since the end of the Little Ice Age.

Recent satellite observations have shown that globally, the permanently oligotrophic subtropical gyres are expanding at a rate of 1% to 4% per year, generally commensurate with global decreases in net primary productivity<sup>9</sup>. In contrast to the global trend, primary production in the NPSG (Fig. 1) has actually increased in recent decades, even as surface waters have become more nutrient-limited<sup>2,3</sup> (Extended Data Fig. 1; Extended Data Table 1). Nitrogen ( $\text{N}_2$ )-fixing prokaryotes, which use the inexhaustible supply of dissolved  $\text{N}_2$  in surface waters, are a key factor in this apparent paradox<sup>4</sup>. This increase in productivity has been accompanied by a ‘domain shift’ from a dominantly eukaryotic to dominantly prokaryotic plankton community<sup>2</sup>. Such ecosystem shifts, with their impact on the export of nutrients to the deep ocean<sup>10</sup> and their sensitivity to past and future climate change, represent major unresolved problems in ocean and global biogeochemical cycles<sup>11,12</sup>.

Oceanographic monitoring at station ALOHA ( $22^\circ 45' \text{W}$ ,  $158^\circ \text{W}$ ; Fig. 1) in the NPSG has suggested a new theory about the dynamic nature of subtropical ocean gyres on seasonal to decadal timescales. Although the NPSG was once considered static and largely unresponsive to climate forcing, it is now apparent that its physical and biological oceanographic variability may be coupled to cyclical climate phenomena such as the North Pacific Gyre Oscillation<sup>3,13</sup>. Therefore, it is not clear whether recent observed changes in subtropical gyre areal extent or the plankton community domain shift at station ALOHA are the result of cyclic, or anomalous (global warming) climate forcing<sup>5</sup>. It is thus imperative to understand these processes over longer timescales than are currently available from the observational record. However,



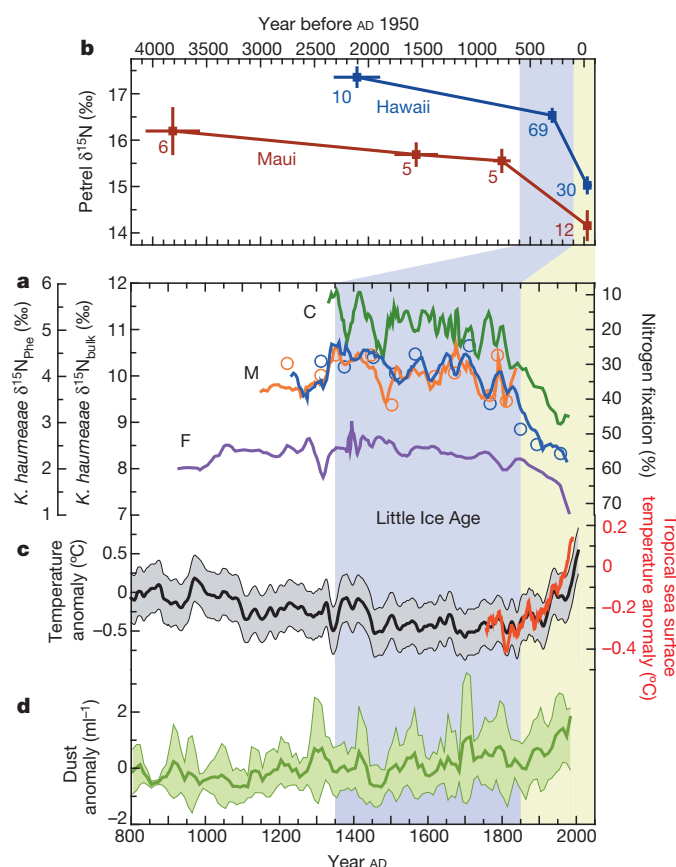
**Figure 1 | NPSG circulation and  $\text{N}^*$  distribution with sample locations.** Contours show climatological mean dynamic height (in units of  $\text{m}^2 \text{s}^{-2}$  relative to the 1,000-dbar level) with arrows showing direction of geostrophic flow. The anticyclonic NPSG lies approximately west of the  $1.6 \text{ m}^2 \text{s}^{-2}$  contour, between  $10^\circ \text{N}$  and  $40^\circ \text{N}$ . The colour scale shows the distribution of  $\text{N}^*$  ( $\text{N}^* = \text{N} - 16\text{P} + 2.9 \mu\text{mol kg}^{-1}$ , where P is phosphorus) (ref. 30). Positive  $\text{N}^*$  indicates where the oceanic N inventory is increased by  $\text{N}_2$  fixation; negative  $\text{N}^*$  indicates where nitrogen is lost to water column denitrification. The lower panel shows *K. haumea* sampling locations—Cross Seamount (C), Makapuu (M) and French Frigate Shoals (F)—and the location of oceanographic station ALOHA (A).

traditional palaeo-archives from sediment cores cannot provide meaningful longer-term context, because the entire Holocene history of the NPSG is represented in less than 10 cm of bioturbated sediment.

Here, we use the unique geochemical time-series data encoded in the skeletons of extraordinarily long-lived, deep-sea corals to reconstruct the first detailed biogeochemical proxy records of the NPSG. The Hawaiian gold coral *Kulamanamana haumea* Sinniger *et al.*, 2013 (previously known as *Gerardia* sp.), secretes a proteinaceous skeleton synthesized from its primary food source: recently exported sinking particles<sup>14</sup>. These growth-layered, decay-resistant skeletons therefore record the biogeochemical signatures of exported production, in a manner analogous to ‘living sediment traps’, with annual- to decadal-scale resolution over millennial timescales<sup>14,15</sup>.

We have generated records of skeletal bulk  $\delta^{15}\text{N}$  ( $\delta^{15}\text{N}_{\text{bulk}}$ ) from specimens of *K. haumea* collected from three sites in the Hawaiian archipelago (Fig. 1). These records all show that from around 1000 AD to 1850 AD,  $\delta^{15}\text{N}_{\text{bulk}}$  showed no long-term, secular trend (Fig. 2a). Then, beginning around the end of the Little Ice Age (around 1850 AD),  $\delta^{15}\text{N}_{\text{bulk}}$

<sup>1</sup>Ocean Sciences Department, University of California, Santa Cruz, California 95064, USA. <sup>2</sup>Lawrence Livermore National Laboratory, Livermore, California 94550, USA. <sup>3</sup>Institute for Marine Sciences, University of California, Santa Cruz, California 95064, USA. <sup>†</sup>Present address: Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado 80309, USA.



**Figure 2 | NPSG  $\delta^{15}\text{N}$  proxy records and their relationship to climate change.** **a**, Records of *K. haumeae*  $\delta^{15}\text{N}_{\text{bulk}}$  (solid lines show the three-point running average, analytical error  $<0.15\text{‰}$ ) and  $\delta^{15}\text{N}_{\text{phe}}$  (corresponding orange and blue circles, analytical error  $<1.0\text{‰}$ ). Abbreviations (M, C, F) as in Fig. 1. Percentage  $\text{N}_2$ -fixation based on mixing between  $\text{NO}_3^-$  and  $\text{N}_2$ -fixation shown on separate axis. **b**, Petrel bone collagen  $\delta^{15}\text{N}$  records<sup>17</sup>. For clarity, records are shown for Hawaii and Maui breeding populations only. Error bars are 1 s.e.m., with sample size indicated. **c**, Northern Hemisphere temperature reconstruction<sup>8</sup> (black line with 95% confidence envelope) and tropical Pacific sea surface temperature reconstruction<sup>7</sup> (red line). **d**, Composite dust record from Himalayan ice cores<sup>29</sup> (green line with 95% confidence envelope).

began to decrease dramatically and monotonically to levels not seen at any point over the mid- to late Holocene epoch<sup>16</sup>. The timing and magnitude of the decrease in  $\delta^{15}\text{N}_{\text{bulk}}$  is similar to recently reported records of  $\delta^{15}\text{N}$  preserved in the bone collagen of Hawaiian petrels (*Pterodroma sandwichensis*)<sup>17</sup> (Fig. 2b), suggesting that the observed changes in  $\delta^{15}\text{N}$  are linked across multiple trophic levels. The most recent parts of these records overlap with instrumental data, over which time the NPSG has become increasingly nutrient-limited and more favourable to  $\text{N}_2$ -fixing diazotrophs (Extended Data Fig. 1; Extended Data Table 1).

Values of  $\delta^{15}\text{N}_{\text{bulk}}$  can be difficult to interpret directly: bulk values represent the effects of baseline nitrogen-source signatures combined with subsequent alterations due to trophic transfer<sup>18</sup>. Recently, analysis of the  $\delta^{15}\text{N}$  of individual amino acids ( $\delta^{15}\text{N}_{\text{aa}}$ ) has emerged as a powerful new tool with which to decouple and unambiguously examine these effects independently of one other<sup>15,18–20</sup>. In heterotrophic organisms, amino acids fall into two isotopically distinct groups<sup>18,21</sup> (Extended Data Fig. 2). One group, the ‘source’ amino acids (SrcAA in Table 1), remain little changed through successive trophic levels of a food web. In contrast, ‘trophic’ amino acids (TrAA in Table 1) become significantly enriched in  $^{15}\text{N}$  with each successive trophic transfer.

To distinguish between the effects of baseline nitrogen-source signatures from subsequent trophic alterations, we measured  $\delta^{15}\text{N}_{\text{aa}}$  on a subset of the samples spanning the bulk records. The values for individual

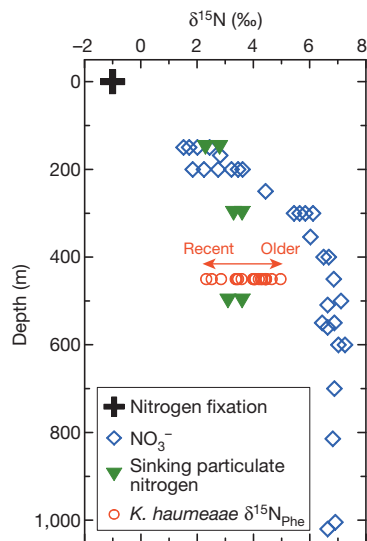
**Table 1 | Correlation of  $\delta^{15}\text{N}_{\text{aa}}$  parameters with  $\delta^{15}\text{N}_{\text{bulk}}$**

Parameter		Slope	Standard error	$r^2$	P value
SrcAA	Gly	0.99	0.31	0.36	<b>0.005</b>
	Ser	1.10	0.34	0.36	<b>0.005</b>
	Lys	0.88	0.19	0.56	<b>&lt;0.001</b>
	Tyr	0.76	0.31	0.27	<b>0.026</b>
	Phe	1.06	0.13	0.80	<b>&lt;0.001</b>
	Average SrcAA	1.00	0.13	0.78	<b>&lt;0.001</b>
TrAA	Glx	0.92	0.43	0.20	<b>0.049</b>
	Asx	0.76	0.26	0.32	<b>0.009</b>
	Ala	0.97	0.42	0.23	<b>0.034</b>
	Ile	1.37	0.37	0.44	<b>0.002</b>
	Leu	0.94	0.33	0.30	<b>0.012</b>
	Pro	0.80	0.29	0.29	<b>0.014</b>
	Val	0.51	0.32	0.13	0.126
	Average TrAA	0.88	0.23	0.46	<b>0.001</b>
Trophic position (average TrAA minus average SrcAA)		-0.01	0.03	0.00	0.880
Trophic position (Glu minus Phe method)		-0.04	0.06	0.02	0.545
$\Sigma V$		0.04	0.12	0.01	0.724

Boldface indicates significant correlations at the  $P < 0.05$  level. Strong correlations are seen for individual source amino acids (SrcAA) and trophic amino acids (TrAA), including overall SrcAA and TrAA group averages. Estimates of trophic position, calculated using two different formulations, as well as microbial resynthesis (as indicated by the  $\Sigma V$  parameter) were uncorrelated with  $\delta^{15}\text{N}_{\text{bulk}}$  ( $n = 20$ ). This indicates that trends in  $\delta^{15}\text{N}_{\text{bulk}}$  timeseries are driven by  $\delta^{15}\text{N}$  at the base of the foodweb. Source amino acids are glycine (Gly), serine (Ser), lysine (Lys), tyrosine (Tyr) and phenylalanine (Phe). Trophic amino acids are glutamic acid + glutamine (Glx), aspartic acid + asparagine (Asx), alanine (Ala), isoleucine (Ile), leucine (Leu), proline (Pro) and valine (Val). The two relative trophic position formulations were:  $\{[(\delta^{15}\text{N}_{\text{average TrAA}} - \delta^{15}\text{N}_{\text{average SrcAA}}) - 3.4]/7.6\} + 1$  (ref. 19) and  $\{[(\delta^{15}\text{N}_{\text{Glu}} - \delta^{15}\text{N}_{\text{Phe}}) - 3.4]/7.6\} + 1$  (ref. 19).  $\Sigma V = 1/7 \Sigma \text{Abs}(\delta^{15}\text{N}_{\text{TrAA}} - \delta^{15}\text{N}_{\text{average TrAA}})$  (ref. 19).

amino acids, as well as SrcAA and TrAA group averages, were significantly correlated with  $\delta^{15}\text{N}_{\text{bulk}}$  with a slope near unity (Table 1). In contrast, calculated trophic position remained essentially constant, and had no statistical relationship with  $\delta^{15}\text{N}_{\text{bulk}}$ . This implies that most of the variability in  $\delta^{15}\text{N}_{\text{bulk}}$  (Fig. 2a) can be attributed to changes in source nitrogen at the base of the foodweb. Among the SrcAA group, phenylalanine (Phe) best preserves the baseline  $\delta^{15}\text{N}$  values, with negligible fractionation through trophic transfers in planktonic food webs<sup>18,21</sup>. In the Makapuu corals (Fig. 1),  $\delta^{15}\text{N}_{\text{phe}}$  decreased from an average of  $4.1 \pm 0.4\text{‰}$  (1 s.d.,  $n = 16$ ) during the pre-1850 period, to a low of  $2.3\text{‰}$  in the most recent part of the record (Fig. 2a). This latter value is directly within the range of present-day thermocline  $\text{NO}_3^-$  ( $\delta^{15}\text{N}_{\text{NO}_3} = 1.5\text{--}2.4\text{‰}$ ,  $n = 4$ ) and sinking particulate nitrogen ( $\delta^{15}\text{N}_{\text{pNsink}} = 2.3\text{--}3.6\text{‰}$ ,  $n = 6$ ) at station ALOHA<sup>22</sup> (Fig. 3), confirming that  $\delta^{15}\text{N}_{\text{phe}}$  in these corals represents a close proxy for baseline  $\delta^{15}\text{N}$  values (Methods). Further, both the  $\delta^{15}\text{N}_{\text{aa}}$  patterns and the analysis of the  $\delta^{15}\text{N}$  of TrAA confirm that neither trophic position nor the microbial resynthesis of sinking particles has significantly affected the observed trends in  $\delta^{15}\text{N}_{\text{bulk}}$  (Table 1, Extended Data Fig. 3).

On the multi-decadal timescale of the observed trends (Fig. 2a), exported productivity in the NPSG is supported by two isotopically distinct nitrogen sources: fixation of dissolved  $\text{N}_2$  ( $\delta^{15}\text{N}_{\text{Nfix}} = -1\text{‰}$ ), and upward transport of  $\text{NO}_3^-$  ( $\delta^{15}\text{N}_{\text{NO}_3} = 6.5\text{‰}$ )<sup>4,22,23</sup>. Assuming that inputs of N must be balanced by exports of sinking particulate nitrogen, the proportional contribution of each source is reflected in  $\delta^{15}\text{N}_{\text{pNsink}}$ <sup>4,22,23</sup>. Therefore, a well-established, two-endmember mixing model— $F_{\text{Nfix}} = 1 - [(\delta^{15}\text{N}_{\text{pNsink}} - \delta^{15}\text{N}_{\text{Nfix}})/(\delta^{15}\text{N}_{\text{NO}_3} - \delta^{15}\text{N}_{\text{Nfix}})]$ , where  $F_{\text{Nfix}}$  is the fraction of  $\text{N}_2$  fixation—has been widely applied to estimate the contribution of  $\text{N}_2$  fixation to export production at station ALOHA (that is, about half at present)<sup>4,10,13,22,23</sup>. By substituting coral  $\delta^{15}\text{N}$  as a proxy for  $\delta^{15}\text{N}_{\text{pNsink}}$ , the long-term change in  $\text{N}_2$  fixation may be estimated. Using this simple model, the change observed in the Makapuu and Cross Seamount records ( $\Delta\delta^{15}\text{N}_{\text{bulk}} = \Delta\delta^{15}\text{N}_{\text{phe}} = -2\text{‰}$ ) indicates a 27% increase in  $\text{N}_2$  fixation since about 1850 AD. For French Frigate Shoals ( $\Delta\delta^{15}\text{N}_{\text{bulk}} = -1.3\text{‰}$ ) the same calculation indicates a 17% increase. We note that the addition or modification of nitrogen-isotope endmembers could potentially affect this interpretation; however,



**Figure 3 | Depth variations in water column  $\delta^{15}\text{N}$  confirm *K. haumea*  $\delta^{15}\text{N}_{\text{Phe}}$  as a proxy for export production.** Particulate nitrogen export at station ALOHA is supported by  $\text{N}_2$  fixation and the upward flux of deep ( $>400$  m)  $\text{NO}_3^-$ . Remineralization of sinking particulate nitrogen contributes to formation of a shallow (150–300 m) pool of isotopically light ( $\sim 2.5\text{‰}$ )  $\text{NO}_3^-$ , which turns over in less than 15 years (ref. 22). Makapuu *K. haumea*  $\delta^{15}\text{N}_{\text{Phe}}$  (analytical error  $<1.0\text{‰}$ ) decreased from an average of  $4.2\text{‰}$  before about 1850, to a low of  $2.3\text{‰}$  in the late twentieth century (red arrow), within range of the shallow  $\delta^{15}\text{N}_{\text{NO}_3}$  and sinking particles<sup>22</sup>.

none of the plausible mechanisms, such as atmospheric or land-based sources of nitrogen or changes in eastern Pacific denitrification, can explain the observed trends (Methods).

We offer two hypotheses for the long-term increase in  $\text{N}_2$  fixation indicated by the *K. haumea* records. The first involves a sustained expansion of the NPSG since the end of the Little Ice Age. This idea is supported by current seasonal trends in the NPSG extent and  $\text{N}_2$  fixation, as well as the  $\delta^{15}\text{N}_{\text{bulk}}$  offsets between sampling locations with respect to the distribution of the nitrogen-to-phosphorus (P) stoichiometric imbalance parameter  $\text{N}^*$  (where  $\text{N}^* = \text{N} - 16\text{P} + 2.9 \mu\text{mol kg}^{-1}$ ) (ref. 30) (Fig. 1). Currently, seasonal fluctuations in gyre areal extent and nutrient inventories are associated with predictable changes in both  $\text{N}_2$  fixation<sup>13,24</sup> and the  $\delta^{15}\text{N}$  of export production: at station ALOHA, the larger summertime gyre corresponds with an approximately  $1.5\text{‰}$  decrease in  $\delta^{15}\text{N}_{\text{pN}_{\text{sink}}}$  (refs 10, 23). Such seasonal changes provide a model for how  $\text{N}_2$  fixation and  $\delta^{15}\text{N}_{\text{pN}_{\text{sink}}}$  might change with progressive gyre expansion over a longer timescale. Further, the relative offsets in the  $\delta^{15}\text{N}_{\text{bulk}}$  records between our three locations, as well as the relative gradients of change observed (Fig. 2a), are also consistent with an expanding gyre margin. Finally, longer-term expansion of NPSG on multiple timescales has also been inferred from satellite imagery<sup>5,9,25</sup>, models<sup>5</sup>, and direct observation records<sup>2,3</sup> (Extended Data Fig. 1; Extended Data Table 1). Although recent shifts could also be related to interannual to decadal natural cycles<sup>3,5,13</sup>, a number of centennial-scale trends, including decreased precipitation<sup>26</sup>, increased tropical Pacific sea surface temperature<sup>7</sup> (Fig. 2b) and inferred shifts in the latitudinal position of the Pacific intertropical convergence zone after the Little Ice Age<sup>27</sup>, are all consistent with the idea of NPSG expansion.

A second hypothesis involves an increase in  $\text{N}_2$  fixation linked to the supply of iron-bearing dust aerosols<sup>11</sup>. Hawaii is significantly affected by bioavailable iron dust originating from Asia, which relaxes the iron limitation for  $\text{N}_2$ -fixing diazotrophs<sup>28</sup>. Millennial-length ice-core records from Himalayan glaciers document up to fourfold increases in dust concentrations over the past century, reflecting warming trends in Asia<sup>29</sup> (Fig. 2c). Such an increase might also account for the trend towards higher rates of  $\text{N}_2$  fixation indicated by the *K. haumea*  $\delta^{15}\text{N}$  records. However, neither gyre expansion or increasing dust deposition are

mutually exclusive hypotheses, and we note that ultimately both are the result of a common driver—the Northern Hemisphere warming trends since the end of the Little Ice Age<sup>8</sup>.

Overall, these data provide strong evidence for a progressive, 17%–27% increase in  $\text{N}_2$  fixation in the NPSG since the mid-nineteenth century, unique with respect to the mid- to late Holocene and comparable in magnitude to the total change in global mean sedimentary  $\delta^{15}\text{N}$  across the Pleistocene–Holocene transition<sup>6</sup>. From a consideration of the mechanisms that can most plausibly explain this shift, we conclude that export productivity in the NPSG has been responding to anomalous climate forcing over the past 150 years or so.

Given that the *K. haumea* records are a proxy for exported productivity, these data indicate that the biogeochemical impacts of changing surface  $\text{N}_2$  fixation propagate at least into the mesopelagic ocean on centennial timescales. This signal, in both timing and amplitude, is propagated into multiple higher trophic levels of the marine food web and is observed in petrel bone bulk  $\delta^{15}\text{N}$  data: that is, the petrel data are probably not reflecting declining trophic position<sup>17</sup>, but instead changes at the base of the food web. These dramatic changes in the world's largest contiguous biome (NPSG) highlight the important role of nitrogen fixation in the response of marine ecosystems to long-term climate change.

## METHODS SUMMARY

Colonies of Hawaiian gold coral (*K. haumea*) were collected from water depths of approximately 450 m with the HURL/NOAA *Pisces V* submersible between 1997 and 2004 (ref. 14 and Extended Data Fig. 4). Air-dried colonies were sectioned near the skeletal base, polished and photographed under a binocular microscope. A computerized Merchantek micromill was used to mill samples, parallel to growth banding, at increments of 0.1 mm, along radial transects from the outer edge to the centre of each section (Extended Data Fig. 5). Each sample represents from 1 to 20 (average 5) years of growth, depending on growth rate. Radiocarbon dating of sample aliquots (Extended Data Table 2) was used for age modelling (Extended Data Fig. 6).

Bulk  $\delta^{15}\text{N}$  isotope ratios—defined as  $\delta^{15}\text{N} = [({}^{15}\text{N}/{}^{14}\text{N})_{\text{sample}}/({}^{15}\text{N}/{}^{14}\text{N})_{\text{standard}} - 1] \times 1,000$ —were measured using a Carlo Erba 1108 elemental analyser interfaced to a ThermoFinnigan Delta Plus XP isotope ratio mass spectrometer (IRMS). Values are reported in per mil (‰) units relative to atmospheric  $\text{N}_2$  ( $\delta^{15}\text{N}_{\text{air}} = 0\text{‰}$ ). Reproducibility, as measured by the difference in sample replicates, averaged  $<0.15\text{‰}$ .

Amino-acid-specific  $\delta^{15}\text{N}_{\text{aa}}$  was measured on sample composites (combining 3 to 10 separate samples to obtain a total mass of 15–20 mg). Composites were hydrolysed in 100 ml of 6 N HCl for 20 h, and spiked with a norleucine internal standard followed by formation of trifluoroacetyl/isopropyl ester derivatives<sup>19</sup>. These derivatives were analysed on a Thermo Trace Ultra gas chromatograph, fitted with a SGE BPX5 capillary column (60 m  $\times$  0.32 mm internal diameter, 1  $\mu\text{m}$  film thickness), in line with the oxidation and reduction furnaces, and linked to a ThermoFinnigan Delta Plus XP IRMS. Samples were analysed in quadruplicate. Analytical accuracy was monitored by analysis of the norleucine internal standard and co-derivatized amino-acid external standards for which authentic  $\delta^{15}\text{N}$  values of each amino acid were determined offline. Reproducibility for individual amino-acid values was typically better than  $1\text{‰}$ .

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 February; accepted 2 October 2013.

Published online 15 December 2013.

- Emerson, S. *et al.* Experimental determination of the organic carbon flux from open-ocean surface waters. *Nature* **389**, 951–954 (1997).
- Karl, D. M., Bidigare, R. R. & Letelier, R. M. Long-term changes in plankton community structure and productivity in the North Pacific Subtropical Gyre: the domain shift hypothesis. *Deep Sea Res. Part II* **48**, 1449–1470 (2001).
- Corno, G. *et al.* Impact of climate forcing on ecosystem processes in the North Pacific Subtropical Gyre. *J. Geophys. Res.* **112**, C04021 (2007).
- Karl, D. *et al.* The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* **388**, 533–538 (1997).
- Henson, S. A. *et al.* Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity. *Biogeosciences* **7**, 621–640 (2010).
- Galbraith, E. D. *et al.* The acceleration of oceanic denitrification during deglacial warming. *Nature Geosci.* **6**, 579–584 (2013).
- Wilson, R. *et al.* Two-hundred-fifty years of reconstructed and modeled tropical temperatures. *J. Geophys. Res.* **111**, C10007 (2006).



8. Mann, M. E. *et al.* Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia. *Proc. Natl Acad. Sci. USA* **105**, 13252–13257 (2008).
9. Polovina, J. J., Howell, E. A. & Abecassis, M. Ocean's least productive waters are expanding. *Geophys. Res. Lett.* **35**, L03618 (2008).
10. Karl, D. M., Church, M. J., Dore, J. E., Letelier, R. M. & Mahaffey, C. Predictable and efficient carbon sequestration in the North Pacific Ocean supported by symbiotic nitrogen fixation. *Proc. Natl Acad. Sci. USA* **109**, 1842–1849 (2012).
11. Falkowski, P. G., Barber, R. T. & Smetacek, V. Biogeochemical controls and feedbacks on ocean primary production. *Science* **281**, 200–206 (1998).
12. Chavez, F. P., Messie, M. & Pennington, J. T. Marine primary production in relation to climate variability and change. *Annu. Rev. Mar. Sci.* **3**, 227–260 (2011).
13. Dave, A. C. & Lozier, M. S. Local stratification control of marine productivity in the subtropical North Pacific. *J. Geophys. Res.* **115**, C12032 (2010).
14. Roark, E. B., Guilderson, T. P., Dunbar, R. B., Fallon, S. J. & Mucciarone, D. A. Extreme longevity in proteinaceous deep-sea corals. *Proc. Natl Acad. Sci. USA* **106**, 5204–5208 (2009).
15. Sherwood, O. A., Lehmann, M. F., Schubert, C. J., Scott, D. B. & McCarthy, M. D. Nutrient regime shift in the western North Atlantic indicated by compound-specific  $\delta^{15}\text{N}$  of deep-sea gorgonian corals. *Proc. Natl Acad. Sci. USA* **108**, 1011–1015 (2011).
16. Guilderson, T. P., McCarthy, M. D., Dunbar, R. B., Englebrecht, A. & Roark, E. B. Late Holocene variations in Pacific surface circulation and biogeochemistry inferred from proteinaceous deep-sea corals. *Biogeosciences* **10**, 6019–6028 (2013).
17. Wiley, A. E. *et al.* Millennial-scale isotope records from a wide-ranging predator show evidence of recent human impact to oceanic food webs. *Proc. Natl Acad. Sci. USA* **110**, 8972–8977 (2013).
18. McClelland, J. W. & Montoya, J. P. Trophic relationships and the nitrogen isotopic composition of amino acids in plankton. *Ecology* **83**, 2173–2180 (2002).
19. McCarthy, M. D., Benner, R., Lee, C. & Fogel, M. L. Amino acid nitrogen isotopic fractionation patterns as indicators of heterotrophy in plankton, particulate, and dissolved organic matter. *Geochim. Cosmochim. Acta* **71**, 4727–4744 (2007).
20. Popp, B. N. *et al.* Insight into the trophic ecology of yellowfin tuna, *Thunnus albacares*, from compound-specific nitrogen isotope analysis of proteinaceous amino acids. *Terr. Ecol.* **1**, 173–190 (2007).
21. Chikaraishi, Y. *et al.* Determination of aquatic food-web structure based on compound-specific nitrogen isotopic composition of amino acids. *Limnol. Oceanogr. Methods* **7**, 740–750 (2009).
22. Casciotti, K. L., Trull, T. W., Glover, D. M., & Davies, D. Constraints on nitrogen cycling at the subtropical North Pacific Station ALOHA from isotopic measurements of nitrate and particulate nitrogen. *Deep Sea Res. Part II* **55**, 1661–1672 (2008).
23. Dore, J. E., Brum, J. R., Tupas, L. M. & Karl, D. M. Seasonal and interannual variability in sources of nitrogen supporting export in the oligotrophic subtropical North Pacific Ocean. *Limnol. Oceanogr.* **47**, 1595–1607 (2002).
24. Church, M. J. *et al.* Physical forcing of nitrogen fixation and diazotroph community structure in the North Pacific subtropical gyre. *Glob. Biogeochem. Cycles* **23**, GB2020 (2009).
25. Behrenfeld, M. J. *et al.* Climate-driven trends in contemporary ocean productivity. *Nature* **444**, 752–755 (2006).
26. Diaz, H. F. & Giambelluca, T. W. Changes in atmospheric circulation patterns associated with high and low rainfall regimes in the Hawaiian Islands region on multiple timescales. *Glob. Planet. Change* **98–99**, 97–108 (2012).
27. Sachs, J. P. *et al.* Southward movement of the Pacific intertropical convergence zone AD 1400–1850. *Nature Geosci.* **2**, 519–525 (2009).
28. Boyle, E. A., Bergquist, B. A., Kayser, R. A. & Mahowald, N. Iron, manganese, and lead at Hawaii Ocean Time-series station ALOHA: temporal variability and an intermediate water hydrothermal plume. *Geochim. Cosmochim. Acta* **69**, 933–952 (2005).
29. Thompson, L. G. *et al.* A high-resolution millennial record of the South Asian monsoon from Himalayan ice cores. *Science* **289**, 1916–1919 (2000).
30. Deutsch, C., Gruber, N., Key, R. M., Sarmiento, J. L. & Ganachaud, A. Denitrification and  $\text{N}_2$  fixation in the Pacific Ocean. *Glob. Biogeochem. Cycles* **15**, 483–506 (2001).

**Acknowledgements** We thank the captain and crew of the RV *Ka'imikai-o-Kanaloa* and the pilots and engineers of the Hawaiian Undersea Research Laboratory's *Pisces V* submersible for their assistance in collecting the specimens presented here. Funding for sample collection was from NOAA/NURP and the National Geographic Society (grant number 7717-04). Radiocarbon analyses were performed under the auspices of the US Department of Energy (grant number DE-AC52-07NA27344). The bulk of the work presented here was funded by the NSF (grant number OCE 1061689).

**Author Contributions** O.A.S. helped conceive the project, prepared samples, performed bulk and compound-specific  $\delta^{15}\text{N}$  analyses and wrote the paper. F.C.B. and J.T.S. assisted in sample preparation and analysis, and commented on the manuscript. T.P.G. and M.D.M. conceived and supervised this project, discussed the results and edited the manuscript.

**Author Information** Data will be digitally archived with the National Oceanic and Atmospheric Administration paleoclimatology datasets (<http://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets>) after publication. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.A.S. ([owen.sherwood@colorado.edu](mailto:owen.sherwood@colorado.edu)).



## METHODS

**Sample collection and preparation.** Colonies of Hawaiian gold coral were collected from water depths of approximately 450 m with the HURL/NOAA *Pisces V* submersible between 1997 and 2004 (ref. 14 and Extended Data Fig. 4). Traditionally known as *Gerardia* sp., the Hawaiian gold coral was recently reclassified as a newly erected genus and species, *Kulumanana haumea*<sup>21</sup>. All of the specimens were collected alive, with the exception of specimen Ger9702 from Makapuu which was dead when collected. Air-dried colonies were sectioned near the skeletal base with a rock saw. Sections 10 mm thick were mounted on 50 mm × 75 mm glass slides with epoxy, ground and polished on diamond laps, ultrasonically cleansed in isopropyl alcohol, and photographed under a binocular microscope. A computerized Merchantek micromill was used to mill samples, parallel to growth banding, at increments of 0.1 mm, along radial transects from the outer edge to the centre of each section (Extended Data Fig. 5). Each approximately 5-mg sample represents from 1 to 20 (average 5) years of growth, depending on growth rate.

**Radiocarbon dating and age models.** Radiocarbon dating was performed on 8 to 10 sample aliquots from each section of *K. haumea*. Radiocarbon measurements were performed at the Center for Accelerator Mass Spectrometry (CAMS), Lawrence Livermore National Laboratory. Aliquots (1 mg) were combusted in individual quartz tubes and reduced to graphite in the presence of iron catalyst. Results include a background and  $\delta^{13}\text{C}$  correction and are reported as Fraction modern  $^{14}\text{C}$  (ref. 32; Extended Data Table 2). Radiocarbon age calibrations and age models were generated with Calam version 2.0 (ref. 33), using the Marine09 database<sup>34</sup> with a local reservoir ( $\Delta R$ ) correction<sup>35</sup> of  $-34 \pm 13$  years, based on a chronology of surface water  $^{14}\text{C}$  from Hawaiian reef corals<sup>36</sup>. Post-bomb values were calibrated directly to the Hawaiian surface water chronology<sup>36</sup>. Age models were fitted with a spline function with a smoothing level of 0.6. Models were run over 1,000 iterations, from which mean ages and 95% confidence levels were calculated (Extended Data Fig. 6).

**Bulk N isotope analysis.**  $\delta^{15}\text{N}$  isotope ratios—defined as  $\delta^{15}\text{N} = [(^{15}\text{N}/^{14}\text{N})_{\text{sample}} / (^{15}\text{N}/^{14}\text{N})_{\text{standard}} - 1] \times 1,000$ —were measured on 0.7-mg aliquots of all the milled samples using a Carlo Erba 1108 elemental analyser interfaced to a ThermoFinnigan Delta Plus XP isotope ratio mass spectrometer (IRMS). Values are reported in per mil (‰) units relative to atmospheric  $\text{N}_2$  ( $\delta^{15}\text{N}_{\text{air}} = 0\text{‰}$ ). Raw isotope values were corrected for instrument drift and linearity effects and calibrated against the in-house isotopic reference materials of the Stable Isotope Lab, University of California, Santa Cruz, using standard laboratory protocols (<http://es.ucsc.edu/~silab/index.php>). Reproducibility, as measured by the difference in sample replicates, averaged  $<0.15\text{‰}$ .

**AA hydrolysis and derivitization.** Sample composites (combining 3 to 10 separate samples to obtain a total mass of 15–20 mg) were hydrolysed in 100 ml of 6 N HCl for 20 h, and spiked with 6  $\mu\text{l}$  of norleucine internal standard ( $\delta^{15}\text{N} = 7.9\text{‰}$ ). Hydrolysates were evaporated to dryness under a stream of  $\text{N}_2$  and stored in a desiccator overnight, followed by formation of trifluoroacetyl/isopropyl ester derivatives<sup>19</sup>.

**Amino-acid compositional analysis.** Amino-acid molar composition was determined with an Agilent 7890A gas chromatograph fitted with a SGE BPX-5 column (60 m × 0.32 mm internal diameter, 1  $\mu\text{m}$  film thickness). Response factors were determined with a dilution series of an external amino-acid standard mixture of 14 common protein amino acids. Reproducibility, as measured by the standard deviation of analytical replicates, averaged  $<5\text{ mol.}\%$ . Arginine (Arg), cysteine (Cys) and histidine (His) concentrations could not be determined owing to their breakdown during acid hydrolysis.

**$\delta^{15}\text{N}$  amino-acid analysis.** Measurement of the  $\delta^{15}\text{N}$  of amino-acid derivatives was performed using previously published procedures<sup>19</sup>. Derivatives were analysed on a Thermo Trace Ultra gas chromatograph, fitted with a SGE BPX5 capillary column (60 m × 0.32 mm internal diameter, 1  $\mu\text{m}$  film thickness), in line with the oxidation and reduction furnaces, and linked to a ThermoFinnigan Delta Plus XP IRMS. Samples were analysed in quadruplicate. Analytical accuracy was monitored by analysis of the norleucine internal standard and co-derivatized amino-acid external standards for which authentic  $\delta^{15}\text{N}$  values of each amino acid were determined offline. Reproducibility for individual amino-acid values was typically better than 1‰. The  $\delta^{15}\text{N}$  of methionine (Met) was not measured owing to insufficient concentrations.

**Oceanographic data.** The dynamic height and  $\text{N}^*$  data in Fig. 1 use gridded data from World Ocean Atlas 2009 ([http://www.nodc.noaa.gov/OC5/WOA09/wao09\\_data.html](http://www.nodc.noaa.gov/OC5/WOA09/wao09_data.html)). Data in Extended Data Fig. 1 and Extended Data Table 1 use ocean station data from World Ocean Database 2009 (<http://www.nodc.noaa.gov/OC5/SELECT/dbsearch/dbsearch.html>) for the region  $17.5\text{--}27.5^\circ\text{N}$ ,  $150\text{--}170^\circ\text{W}$ . Data were filtered for outliers and plotted using Ocean Data View software, version 4.3.10 (R. Schlitzer, <http://odv.awi.de>).

**Climate records.** Temperature record in Fig. 2c is from the NHHAD\_EIV Northern Hemisphere land and ocean multiple proxy temperature reconstruction of ref. 8, smoothed with a 15-year running average. The tropical Pacific sea surface temperature record is from the compilation of reef coral oxygen isotope records of ref. 7, smoothed with a 15-year running average. The dust record in Fig. 2d is from

a compilation of the Dasuopo, Dunde and Guliyu Himalayan ice cores<sup>29</sup>. Individual decadal records were normalized by subtracting the mean and dividing by the standard deviation for each record. The composite record is the average of the three individual records, smoothed with a 20-year running average.

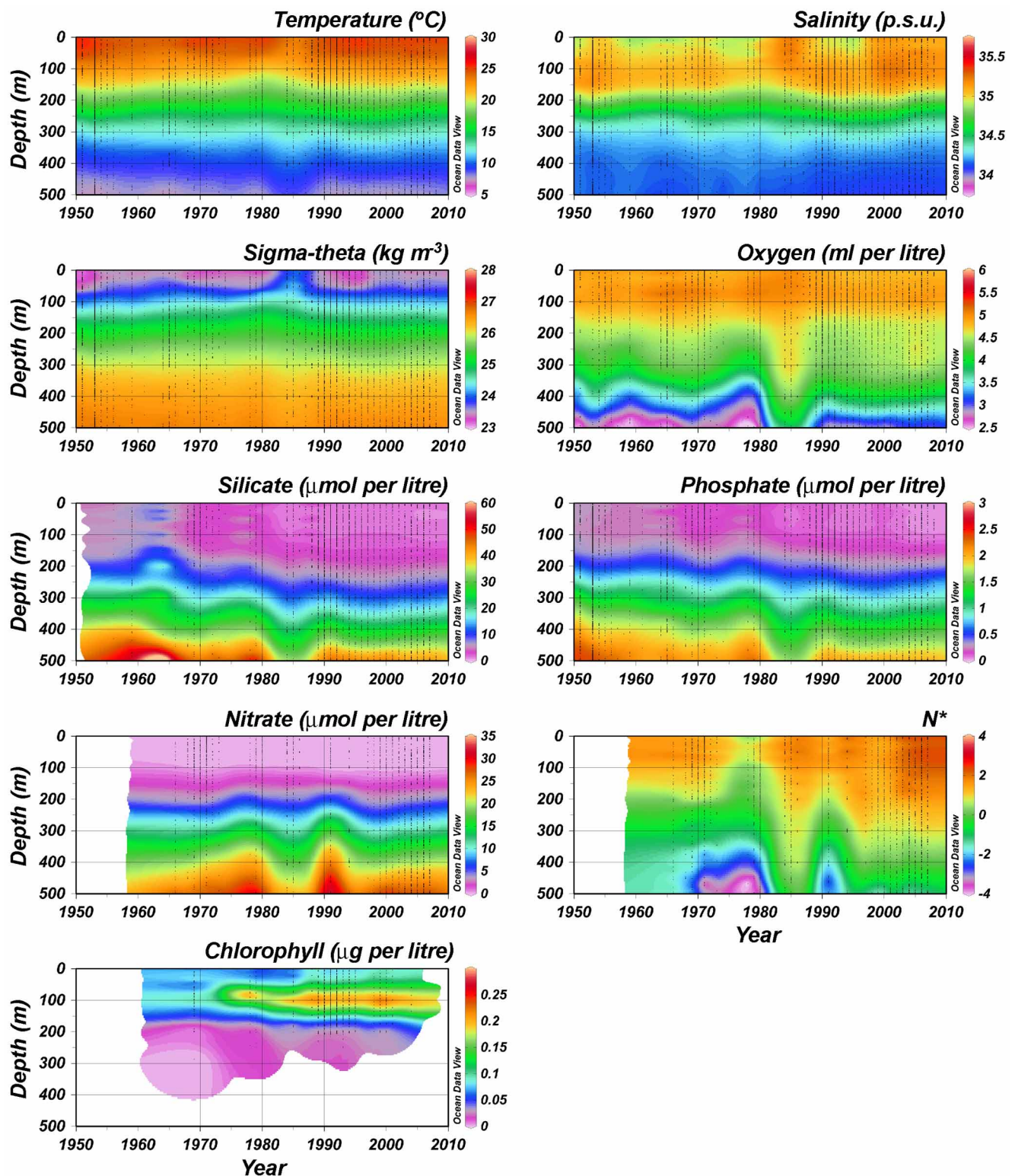
**Deep-sea coral  $\delta^{15}\text{N}$  as a proxy for sinking particulates.** It has been well established that the skeletons of *K. haumea* and other deep-sea proteinaceous corals are biogeochemically tightly coupled to surface waters. The presence of bomb radiocarbon (that is, post-1950) signatures in the living tissues and skeletal protein of deep-sea corals establishes recently exported particles as their main food source<sup>37</sup>. The incorporation of detailed bomb-radiocarbon chronologies within the skeletal growth rings, without any significant attenuation or time-lag relative to surface water chronologies, further establishes that euphotic zone biogeochemical signatures are transmitted to deep-sea corals efficiently and rapidly<sup>14,15</sup>. The  $\delta^{15}\text{N}_{\text{Phe}}$  signature in the Makapuu *K. haumea* specimens closely matches the  $\delta^{15}\text{N}$  of both sinking particulates and thermocline  $\text{NO}_3^-$  (Fig. 3). Data from two other oceanographic regions, the northwest Atlantic and the denitrification-affected eastern north Pacific, show that coral  $\delta^{15}\text{N}_{\text{Phe}}$  values closely follow the same pattern as  $\text{NO}_3^-$  with respect to  $\text{N}^*$  (Extended Data Fig. 7).

**Steady state isotopic mass balance model assumptions.** The isotopic mass balance model assumes that the export flux of particulate N at station ALOHA and surrounding waters of the NSPG is balanced by  $\text{N}_2$  fixation and upward flux of  $\text{NO}_3^-$  (refs 4, 10, 13, 22 and 23). There is no isotopic impact from incomplete  $\text{NO}_3^-$  assimilation, because  $\text{NO}_3^-$  in the mixed layer is exhausted to  $<0.2\text{ }\mu\text{mol kg}^{-1}$  year-round (Hawaii Ocean Time-series Data Organization and Graphical System (HOT-DOGS); see <http://hahana.soest.hawaii.edu/hot/hot-dogs/interface.html>). Additional nitrogen isotopic endmembers are negligible to the nitrogen budget. Atmospheric deposition of organic nitrogen accounts for  $<2\%$  of the particulate nitrogen export at 150 m (refs 4, 23 and 30). Terrestrial runoff of fertilizers (values near  $-2\text{‰}$ ) can depress local seawater  $\delta^{15}\text{N}$ , but the similar trends at all three locations, including French Frigate Shoals (Fig. 1), rules out any direct land-use effect from the Hawaiian islands.

Finally, modification of Pacific midwater  $\delta^{15}\text{N}_{\text{NO}_3}$  could arise from changes in the rates of denitrification along the eastern Pacific margin<sup>38</sup>. However, this is very unlikely to account for the trends in  $\delta^{15}\text{N}$  of Hawaiian *K. haumea* (Fig. 2) for the following reasons. (1) Denitrification would have to decrease to account for the observed downward trends in *K. haumea*  $\delta^{15}\text{N}$ , whereas in fact denitrification has increased in global oxygen minimum zones, including the ETNP<sup>39,40</sup>. (2) The geostrophic circulation isolates the NPSG from a tongue of denitrification-affected waters, of which the northern boundary lies south of the Hawaiian archipelago at a latitude of  $10^\circ\text{N}$  (refs 30 and 41). (3) Oceanographic data from the Hawaiian archipelago show that oxygen levels in the upper 500 m have remained constant or have increased (Extended Data Fig. 1; Extended Data Table 1). This observation is consistent with expansion of the NPSG and inconsistent with any impact of denitrification from the eastern Pacific margin.

- Sinniger, F., Ocana, O. V. & Baco, A. R. Diversity of zoanthids (Anthozoa: Hexacorallia) on Hawaiian seamounts: description of the Hawaiian gold coral and additional zoanthids. *PLoS ONE* **8**, e52607 (2013).
- Stuiver, M. & Polach, H. A. Reporting of  $^{14}\text{C}$  data. *Radiocarbon* **19**, 355–363 (1977).
- Blaauw, M. Methods and code for 'classical' age-modelling of radiocarbon sequences. *Quat. Geochronol.* **5**, 512–518 (2010).
- Reimer, P. J. *et al.* IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon* **51**, 1111–1150 (2009).
- Stuiver, M. & Braziunas, T. F. Modeling atmospheric  $^{14}\text{C}$  influences and  $^{14}\text{C}$  ages of marine samples to 10,000 BC. *Radiocarbon* **35**, 137–189 (1993).
- Druffel, E. R. M. *et al.* Changes of subtropical North Pacific radiocarbon and correlation with climate variability. *Radiocarbon* **43**, 15–25 (2001).
- Griffin, D. W. & Druffel, E. R. M. Sources of carbon to deep-sea corals. *Radiocarbon* **31**, 533–543 (1989).
- Sigman, D., DiFiore, P. J., Hain, M. P., Deutsch, C. & Karl, D. M. Sinking organic matter spreads the nitrogen isotope signal of pelagic denitrification in the North Pacific. *Geophys. Res. Lett.* **36**, L08605 (2009).
- Gutiérrez, D. *et al.* Rapid reorganization in ocean biogeochemistry off Peru towards the end of the Little Ice Age. *Biogeosciences* **6**, 835–848 (2009).
- Stramma, L., Johnson, G. C., Sprintall, J. & Mohrholz, V. Expanding oxygen-minimum zones on the tropical oceans. *Science* **320**, 655–658 (2008).
- Reid, J. L. On the total geostrophic circulation of the Pacific ocean: flow patterns, tracers, and transports. *Prog. Oceanogr.* **39**, 263–352 (1997).
- Karl, D. A sea of change: biogeochemical variability in the North Pacific Subtropical Gyre. *Ecosystems* **2**, 181–214 (1999).
- Hare, E. P., Fogel, M. L., Stafford, T. W., Mitchell, A. D. & Hoering, T. C. The isotopic composition of carbon and nitrogen in individual amino acids isolated from modern and fossil proteins. *J. Archaeol. Sci.* **18**, 277–292 (1991).
- Styring, A. K., Sealy, J. C. & Evershed, R. P. Resolving the bulk  $\delta^{15}\text{N}$  values of ancient human and animal bone collagen via compound-specific nitrogen isotope analysis of constituent amino acids. *Geochim. Cosmochim. Acta* **74**, 241–251 (2010).

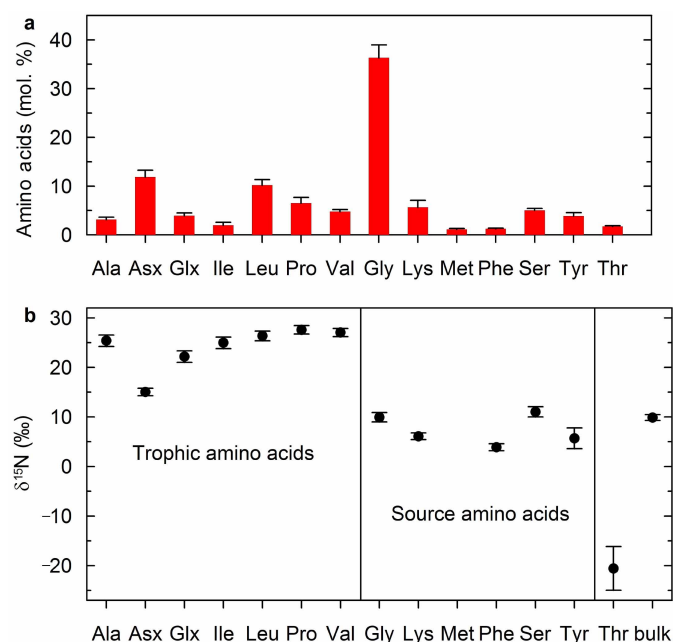
45. Germain, L. R., Koch, P. L., Harvey, J. T. & McCarthy, M. D. Nitrogen isotope fractionation in amino acids from harbor seals: implications for compound-specific trophic position calculations. *Mar. Ecol. Prog. Ser.* **482**, 265–277 (2013).
46. Gruber, N. *The Dynamics of the Marine Nitrogen Cycle and its Influence on Atmospheric CO<sub>2</sub> Variations* NATO Science Series Vol. 40, 97–148 (Springer, 2004).
47. Broek, T. A. B., Walker, B. D., Andreason, D. H. & McCarthy, M. D. High precision measurement of phenylalanine  $\delta^{15}\text{N}$  values for environmental samples: a new approach coupling high-pressure liquid chromatography (HPLC) purification and elemental analysis-isotope ratio mass spectrometry (EA-IRMS). *Rapid Commun. Mass Spectrom.* **27**, 1–11 (2013).



**Extended Data Figure 1 | World Ocean Database 2009 instrumental records demonstrate physical and biological changes in the NPSG since 1950.**

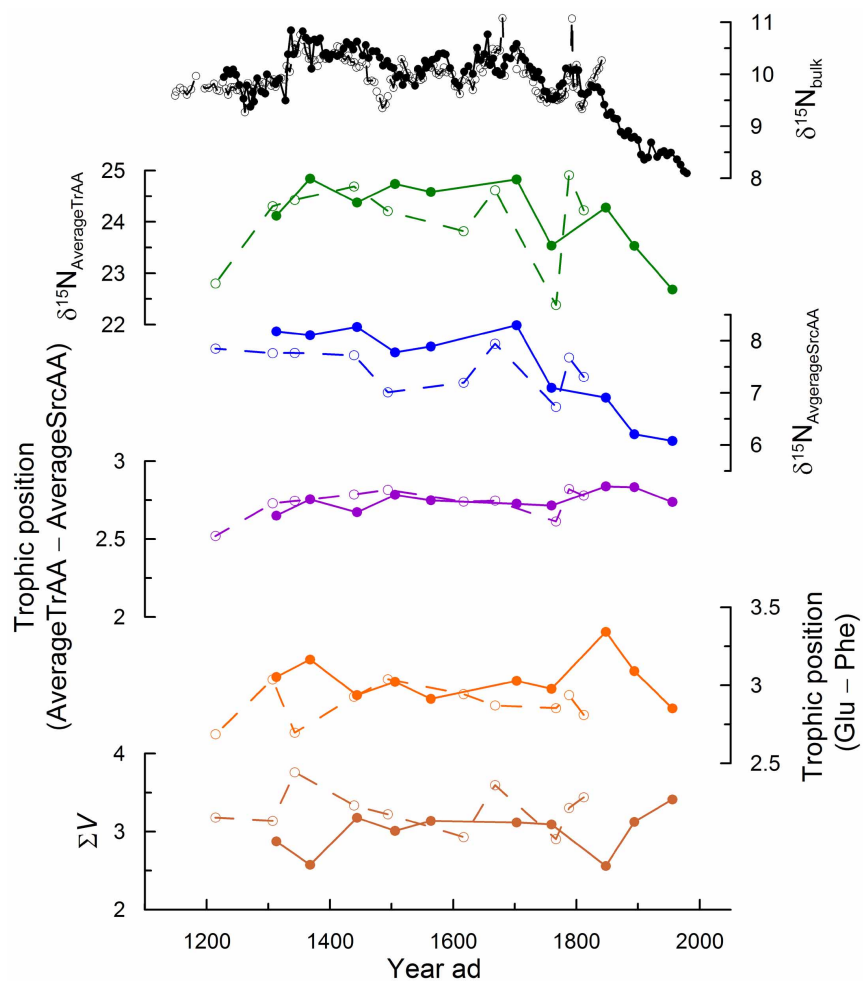
Increasing salinity and temperature is accompanied by a strong decrease in silicate and phosphate nutrient concentrations, and an increase in  $N^*$ . These data are consistent with the previously observed shift from a dominantly

eukaryotic to dominantly prokaryotic ( $N_2$ -fixation) ecosystem in the NPSG<sup>2,3,42</sup>. Increasing rates of change with depth suggest large-scale changes in oceanographic circulation, consistent with expansion of the NPSG. Rates of change across separate depth bins are provided in Extended Data Table 1.

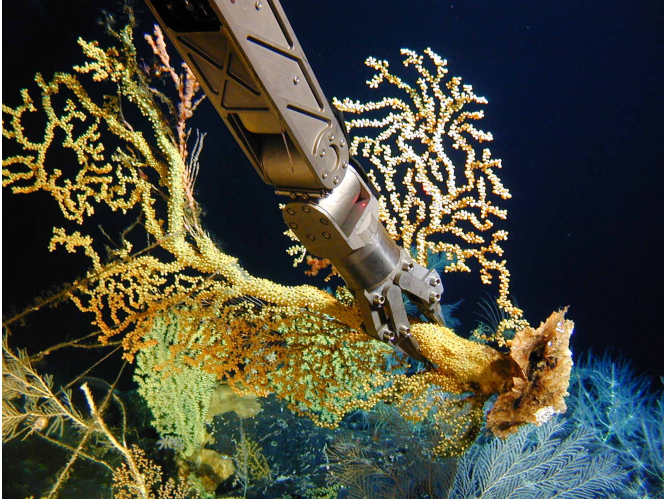


**Extended Data Figure 2 | Distribution of amino-acid concentration and  $\delta^{15}\text{N}_{\text{aa}}$  in *K. haumeeae*.** **a**, Amino-acid mole per cent composition. **b**,  $\delta^{15}\text{N}_{\text{aa}}$  data, showing two main groups: relatively  $\delta^{15}\text{N}$ -enriched 'trophic' amino acids (TrAA), and relatively lower- $\delta^{15}\text{N}$  'source' amino acids (SrcAA). These patterns are very close to those of heterotrophic fresh biomass<sup>18,19,21</sup>. Together with low  $\Sigma V$  values (Table 1), this supports the use of  $\delta^{15}\text{N}_{\text{Phe}}$  as a proxy for the  $\delta^{15}\text{N}$  of exported production, and indicates that  $\delta^{15}\text{N}$  values have not undergone any significant diagenetic alteration. Extremely low values of threonine (Thr) are consistent with previous observations<sup>43,44</sup>, and this amino acid is now understood to be neither a trophic nor source amino acid<sup>45</sup>. Mean  $\delta^{15}\text{N}_{\text{bulk}}$  is shown for context. Error bars represent 1 s.d. ( $n = 20$ ).

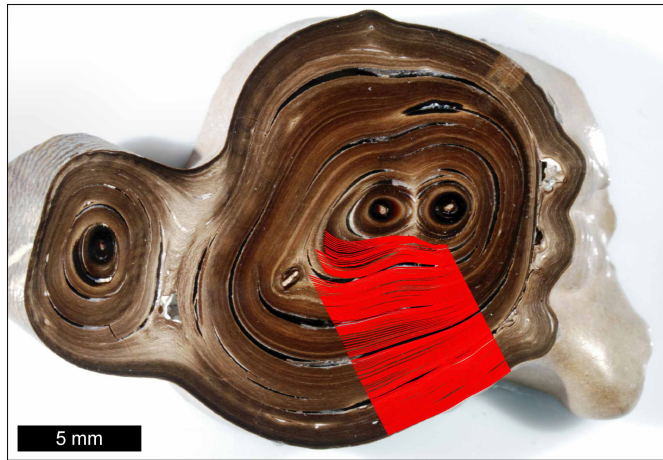




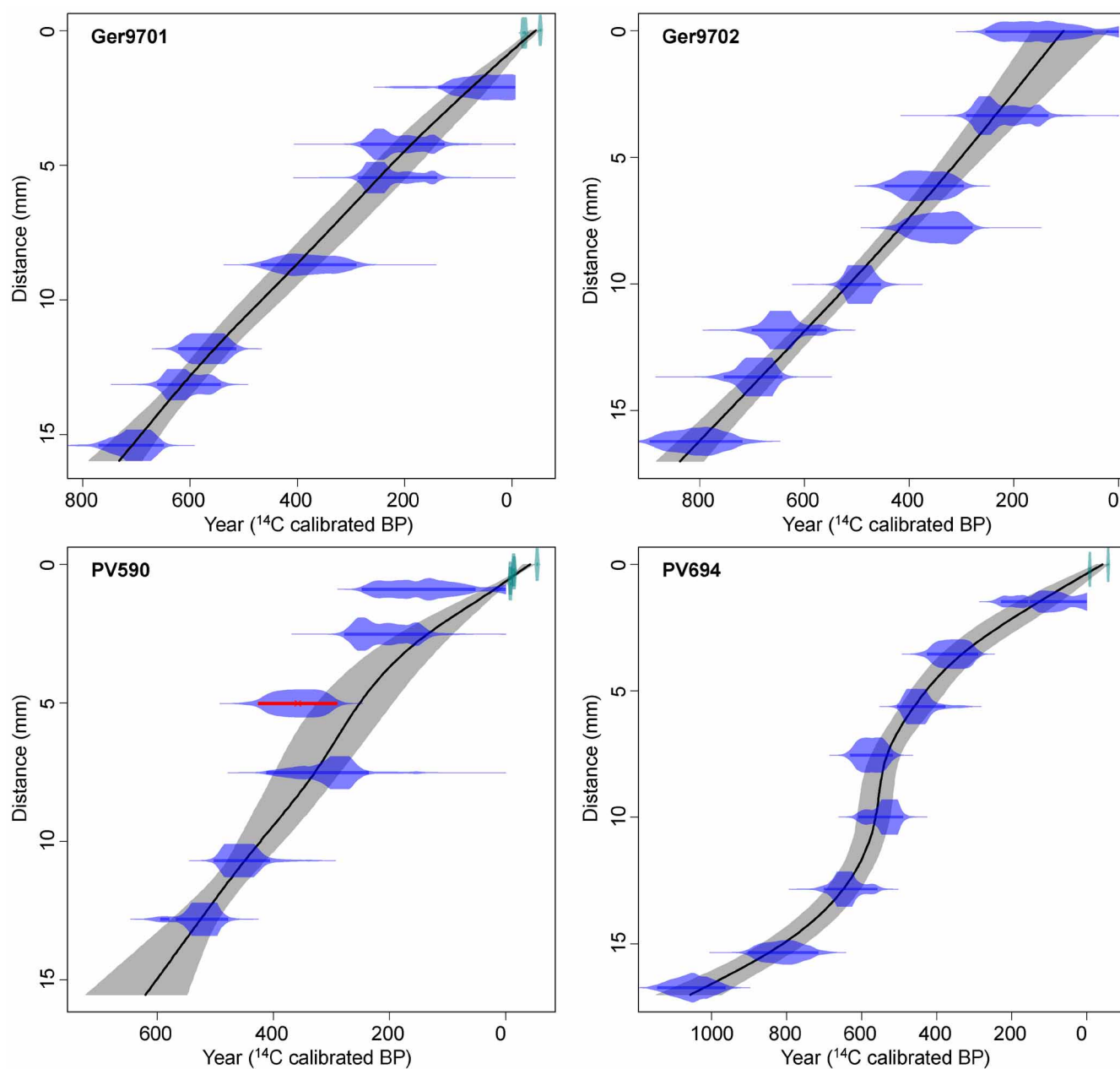
**Extended Data Figure 3 | Timeseries of  $\delta^{15}\text{N}_{\text{aa}}$  parameters in comparison to  $\delta^{15}\text{N}_{\text{bulk}}$  from two *K. haumeae* specimens from Makapuu.** Closed symbols, specimen Ger9701; open symbols, specimen Ger9702. Trophic position and  $\Sigma V$  parameters are defined in Table 1.



**Extended Data Figure 4 | Example *K. haumea* colony photographed *in situ*.** Photo credit: NOAA Hawaiian Undersea Research Laboratory, DSRV Pisces Pilots & Engineers, 2004.



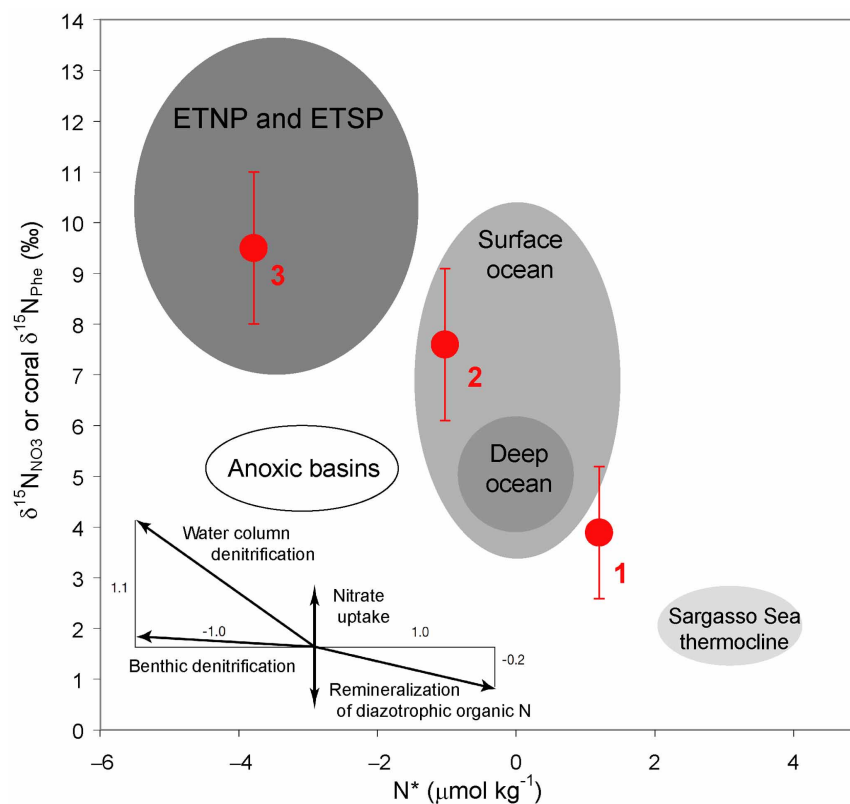
**Extended Data Figure 5 | Example *K. haumea* skeletal cross-section.**  
Specimen Ger9701 from Makapuu. Red lines indicate micromill transect.



**Extended Data Figure 6 | Radiocarbon age models.** Blue-shaded areas represent conventional  $^{14}\text{C}$  calibrated age distributions. Green-shaded areas are post-bomb (post-1950) age distributions. Black curves and grey-shaded regions represent spline fits with 95% confidence intervals, respectively. The average

age model error is 45 years. Outliers (indicated in red) were excluded from age models owing to age reversals. The shape of the curves reflects variable growth rates in the four coral samples illustrated.





**Extended Data Figure 7 | Relationship between  $\delta^{15}\text{N}$  and  $\text{N}^*$  across different oceanographic regions.** The  $\delta^{15}\text{N}_{\text{NO}_3}$  is driven by isotopic fractionation processes of denitrification, nitrate uptake and remineralization of diazotrophic organic nitrogen (ETNP, Eastern Tropical North Pacific; ETSP, Eastern Tropical South Pacific; redrawn from ref. 46). Existing deep-sea coral  $\delta^{15}\text{N}_{\text{Phe}}$  data (red points) closely follow the same overall pattern. Data points are: (1) Hawaii *K. haumea* (2 specimens, 20 measurements;

this paper); (2) Northwest Atlantic *Primnoa resedaeformis* (2 specimens, 8 measurements<sup>15</sup>); (3) Monterey Bay *Isidella* sp. (1 specimen, 10 measurements<sup>47</sup>); Error bars show the total range of measurements. Corresponding values of  $\text{N}^*$  are obtained from the nearest surface water grid point of the World Ocean Atlas 2009 (<http://www.nodc.noaa.gov/OC5/WOA09/wao09data.html>).

Extended Data Table 1 | World Ocean Database 2009 physical and chemical parameters

Parameter	0-150 m		151-300 m		301-500 m	
	Slope $\pm$ Std.Error	P value	Slope $\pm$ Std.Error	P value	Slope $\pm$ Std.Error	P value
Temperature ( $^{\circ}\text{C}$ )	+0.004 $\pm$ 0.007	0.566	+0.017 $\pm$ 0.006	<b>0.009</b>	+0.007 $\pm$ 0.005	0.109
Salinity (p.s.u.)	+0.003 $\pm$ 0.001	<b>0.012</b>	+0.002 $\pm$ 0.001	<b>0.02</b>	−0.001 $\pm$ 0.000	<b>0.005</b>
Sigma-theta ( $\text{kg m}^{-3}$ )	+0.001 $\pm$ 0.003	0.656	−0.002 $\pm$ 0.001	<b>0.031</b>	−0.002 $\pm$ 0.001	<b>0.008</b>
Oxygen (ml per litre)	−0.001 $\pm$ 0.001	0.342	+0.003 $\pm$ 0.002	0.062	+0.013 $\pm$ 0.005	<b>0.016</b>
Silicate ( $\mu\text{mol per litre}$ )	−0.074 $\pm$ 0.023	<b>0.003</b>	−0.121 $\pm$ 0.051	<b>0.022</b>	−0.233 $\pm$ 0.097	<b>0.021</b>
Phosphate ( $\mu\text{mol per litre}$ )	−0.004 $\pm$ 0.001	<b>&lt;0.001</b>	−0.006 $\pm$ 0.001	<b>0.001</b>	−0.008 $\pm$ 0.002	<b>0.001</b>
Nitrate ( $\mu\text{mol per litre}$ )	−0.004 $\pm$ 0.002	0.071	−0.004 $\pm$ 0.029	0.888	−0.007 $\pm$ 0.060	0.910
N*	+0.035 $\pm$ 0.011	<b>0.004</b>	+0.047 $\pm$ 0.014	<b>0.002</b>	+0.064 $\pm$ 0.031	<b>0.054</b>
Chlorophyll ( $\mu\text{g per litre}$ )	+0.002 $\pm$ 0.001	<b>0.014</b>	+0.001 $\pm$ 0.000	<b>0.006</b>	NA	NA

Changes in the parameters in the NPSG are shown from 1950 to 2010, across multiple depth bins. Boldface *P* values indicate statistically significant trends at the  $P < 0.05$  level. Sigma- $\theta$  is the density of sea water at a reference level of 0 dbar (that is, at a depth of 0 m). NA, not available.

Extended Data Table 2 | *K. haumea* radiocarbon data

Coral	Location	CAMS ID	Sample	Distance (mm)	$\delta^{13}\text{C}$	Fraction modern	Error	D <sup>14</sup> C	Error	<sup>14</sup> C Age	Error
Ger9701	Makapuu	156055	1	0.10	-16.9	1.1138	0.0042	113.8	4.2	>Modern	
		156056	20	2.12	-15.7	0.9482	0.0029	-51.8	2.9	425	25
		156057	41	4.24	-15.6	0.9339	0.0035	-66.1	3.5	550	30
		156058	52	5.47	-15.7	0.9321	0.0028	-67.9	2.8	565	25
		156059	81	8.72	-16.0	0.9147	0.0046	-85.3	4.6	715	45
		156060	108	11.84	-16.1	0.8884	0.0028	-111.6	2.8	950	25
		156061	121	13.17	-16.4	0.8816	0.0029	-118.4	2.9	1015	30
		156062	141	15.43	-16.3	0.8684	0.0029	-131.6	2.9	1135	30
Ger9702	Makapuu	153722	4z	0.04	-15.7	0.9414	0.0039	-58.6	3.9	485	35
		153723	40	3.35	-15.6	0.9334	0.0034	-66.6	3.4	555	30
		156064	64	6.12	-15.6	0.9170	0.0030	-83.0	3.0	695	30
		153724	80	7.78	-16.1	0.9204	0.0030	-79.6	3.0	665	30
		156065	102	10.02	-16.4	0.9006	0.0026	-99.4	2.6	840	25
		153725	120	11.82	-16.2	0.8780	0.0037	-122.0	3.7	1045	35
		156066	138	13.67	-16.4	0.8711	0.0029	-128.9	2.9	1110	30
		153726	160	16.21	-16.8	0.8583	0.0037	-141.7	3.7	1225	35
PV590	Cross Seamount	157079	2	0.20	-17.0	1.0736	0.0044	73.6	4.4	>Modern	
		157080	3	0.30	-17.4	1.0855	0.0038	85.5	3.8	>Modern	
		157081	5	0.50	-16.7	0.9714	0.0028	-28.6	2.8	>Modern	
		157082	7	0.70	-16.7	0.9592	0.0028	-40.8	2.8	335	25
		157083	9	0.90	-16.6	0.9419	0.0027	-58.1	2.7	480	25
		157084	25	2.52	-16.5	0.9347	0.0027	-65.3	2.7	540	25
		157085	50	5.02	-16.1	0.9186	0.0027	-81.4	2.7	680	25
		157086	75	7.52	-15.5	0.9264	0.0038	-73.6	3.8	615	35
PV694	French Frigate Shoals	157087	106	10.69	-16.4	0.9073	0.0024	-92.7	2.4	780	25
		157088	127	12.81	-16.3	0.8956	0.0026	-104.4	2.6	885	25
		159298	3	0.20	-16.5	0.9699	0.0031	-30.1	3.1	>Modern	
		159299	15	1.47	-15.7	0.9441	0.0030	-55.9	3.0	460	30
		159300	33	3.55	-15.8	0.9186	0.0028	-81.4	2.8	680	25
		159301	52	5.62	-16.3	0.9082	0.0030	-91.8	3.0	775	30
		159302	71	7.54	-16.1	0.8884	0.0030	-111.6	3.0	950	30
		159303	92	9.99	-16.5	0.8941	0.0029	-105.9	2.9	900	30
		159304	112	12.84	-16.5	0.8779	0.0037	-122.1	3.7	1045	35
		159305	152	15.34	-16.6	0.8579	0.0039	-142.1	3.9	1230	40
		159433	132	16.74	-16.5	0.8326	0.0027	-167.4	2.7	1470	30

Note that specimen Ger9702 was collected dead; the other three specimens were collected alive. CAMS, Center for Accelerator Mass Spectrometry.

# Low investment in sexual reproduction threatens plants adapted to phosphorus limitation

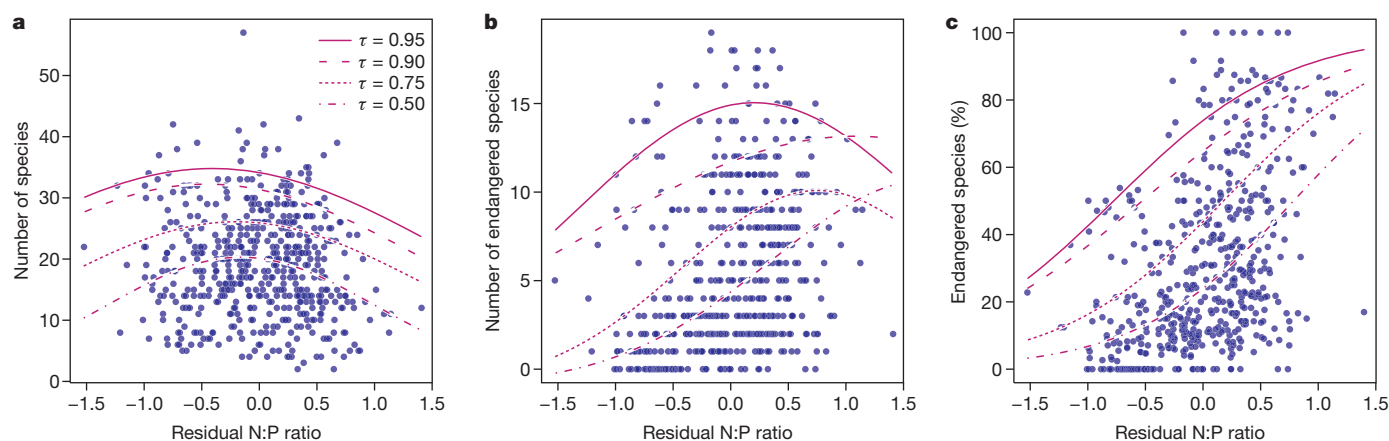
Yuki Fujita<sup>1</sup>, Harry Olde Venterink<sup>2</sup>, Peter M. van Bodegom<sup>3</sup>, Jacob C. Douma<sup>4</sup>, Gerrit W. Heil<sup>5</sup>, Norbert Hölzel<sup>6</sup>, Ewa Jabłońska<sup>7</sup>, Wiktor Kotowski<sup>7</sup>, Tomasz Okruszko<sup>8</sup>, Paweł Pawlikowski<sup>7</sup>, Peter C. de Ruiter<sup>9</sup> & Martin J. Wassen<sup>10</sup>

Plant species diversity in Eurasian wetlands and grasslands depends not only on productivity but also on the relative availability of nutrients, particularly of nitrogen and phosphorus<sup>1–4</sup>. Here we show that the impacts of nitrogen:phosphorus stoichiometry on plant species richness can be explained by selected plant life-history traits, notably by plant investments in growth versus reproduction. In 599 Eurasian sites with herbaceous vegetation we examined the relationship between the local nutrient conditions and community-mean life-history traits. We found that compared with plants in nitrogen-limited communities, plants in phosphorus-limited communities invest little in sexual reproduction (for example, less investment in seed, shorter flowering period, longer lifespan) and have conservative leaf economy traits (that is, a low specific leaf area and a high leaf dry-matter content). Endangered species were more frequent in phosphorus-limited ecosystems and they too invested little in sexual reproduction. The results provide new insight into how plant adaptations to nutrient conditions can drive the distribution of plant species in natural ecosystems and can account for the vulnerability of endangered species.

Species diversity is influenced both by overall nutrient availability and by nutrient stoichiometry—that is, how the ratio of available nutrients relates to their consumers' requirements<sup>5</sup>. In terrestrial plant communities the two nutrients that most frequently limit growth are nitrogen (N) and phosphorus (P)<sup>6</sup>. It has long been recognized that in sites that are N- or P-limited, the species assembly is different<sup>5,6</sup>; this difference is also reflected in the occurrence of endangered species<sup>1–3,7,8</sup>, most of

which are found in P-limited sites. The explanation of why these species are associated with P-limited sites may lie in their functional traits.

In general, fast-growing species dominate in nutrient-rich environments, whereas slow-growing species dominate in nutrient-poor conditions<sup>9,10</sup>. With respect to the N:P stoichiometry, a number of studies indicate that low N:P-ratio environments favour fast-growing species with long roots, or species that fix N, whereas high N:P-ratio environments favour slow-growing species with specialized P uptake traits; for example, cluster roots, arbuscular mycorrhizae or high phosphatase activity<sup>3,11–16</sup>. The association between fast-growing species and low N:P ratios is also consistent with the growth rate hypothesis<sup>17</sup>, which states that a fast growth rate is enabled by high investment in P-rich RNA, resulting in relatively high leaf P concentrations and concomitant low N:P ratios. However, these particular traits do not necessarily explain differences in total species richness along N:P availability gradients and, moreover, it seems possible that the selection for these traits may depend on environmental conditions other than the relative availability of N and P. We looked for an explanation based on inherent plant life-history traits, particularly investments in growth versus reproduction. We had access to a large comparative data set, which enabled us to separate the effects of overall nutrient availability from those of N:P stoichiometry across many species. In addition, by linking our traits analysis to the Red List statuses of the species involved (Red Lists of seven different countries), we were able to assess the mechanisms that might account for why certain species are more vulnerable for extinction than others.

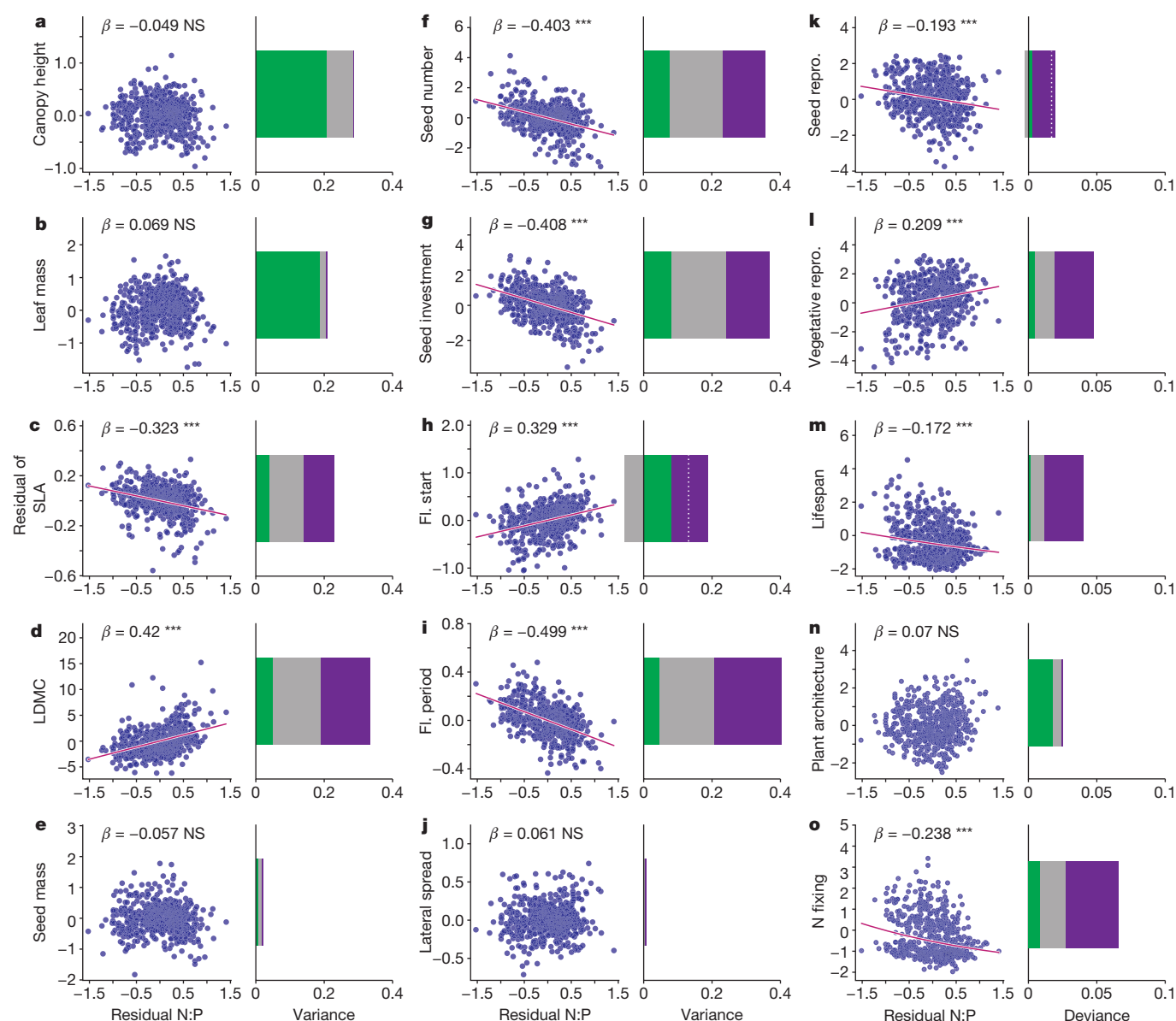


**Figure 1 | Relationship between biodiversity indices of vascular plants and N:P ratio corrected for productivity effects.** a–c, Tested biodiversity indices are the number of species (a), the number of endangered species (b), and the percentage of endangered species (c) (number of sites ( $n$ ) = 539). N:P ratio was corrected for the confounding effects of productivity by using the residual

values of N:P ratio regressed by productivity (see Supplementary Discussion 1 and Extended Data Fig. 1).  $\tau$ th quantile regression functions ( $\tau = 0.50, 0.75, 0.90, 0.95$ ) are also shown. See Extended Data Fig. 2 for the 95% confidence intervals of the regression coefficients. The results were not biased by the selection of habitat types in our data set (Extended Data Fig. 3).

<sup>1</sup>KWR Watercycle Research Institute, PO Box 1072, 3430 BB Nieuwegein, The Netherlands. <sup>2</sup>Plant Biology and Nature Management, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. <sup>3</sup>Department of Ecological Science, VU University Amsterdam, de Boelelaan 1085, 1081 HV Amsterdam, The Netherlands. <sup>4</sup>Centre for Crop Systems Analysis, Wageningen University, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands. <sup>5</sup>Faculty of Science, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands. <sup>6</sup>Institute of Landscape Ecology, Faculty of Geosciences, University of Münster, Robert-Koch-Strasse 28, 48149 Münster, Germany. <sup>7</sup>Department of Plant Ecology and Environmental Conservation, Institute of Botany, University of Warsaw, Al. Ujazdowskie 4, Warsaw, Poland. <sup>8</sup>Warsaw University of Life Sciences, Faculty of Civil and Environmental Engineering, 02-722 Warsaw, Nowoursynowska 166, Poland. <sup>9</sup>Biometris, Wageningen University, PO Box 100, 6700 AC Wageningen, The Netherlands. <sup>10</sup>Environmental Sciences, Copernicus Institute of Sustainable Development, Faculty of Geosciences, Utrecht University, PO Box 80115, 3508 TC Utrecht, The Netherlands.





**Figure 2 | Relationship between N:P ratios and community-mean trait values of herbaceous vascular plant species, after removing confounding effects of productivity.** The residual values of N:P ratio (x axis) and community-mean traits (y axis; Extended Data Fig. 4) regressed by productivity were used (see Supplementary Discussion 1 and Extended Data Fig. 1). **a–o**, The 15 traits tested are canopy height (**a**, number of sites ( $n$ ) = 530), leaf mass (**b**,  $n$  = 525), specific leaf area (**c**,  $n$  = 529), leaf dry matter content (**d**,  $n$  = 525), seed mass (**e**,  $n$  = 533), seed number per shoot (**f**,  $n$  = 524), seed investment (**g**,  $n$  = 523), month flowering started (**h**,  $n$  = 528), flowering period (**i**,  $n$  = 528), lateral spread (**j**,  $n$  = 526), reproduction by seeds (**k**,  $n$  = 528), vegetative reproduction (**l**,  $n$  = 528), lifespan (**m**,  $n$  = 531), plant architecture (**n**,  $n$  = 533), and N fixation (**o**,  $n$  = 502). Extended Data Table 1

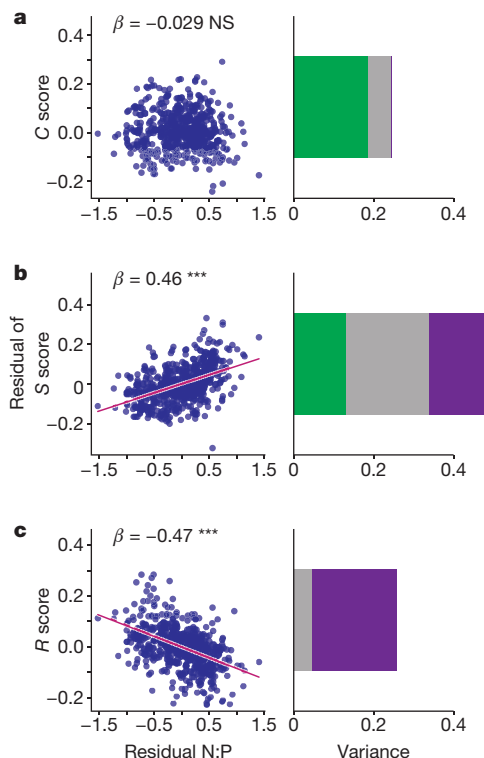
The data set consisted of 599 field plots in herbaceous ecosystems in Eurasia, with data on plant species composition, aboveground productivity of vascular plants as a proxy for overall nutrient availability, and the N:P:potassium (K) ratio in the aboveground plant biomass as a proxy for nutrient stoichiometry (See Supplementary Discussion 1). Sites considered to be K-limited ( $n$  = 60, based on N:K and K:P ratios<sup>1</sup>) were excluded from further analysis. Across the remaining part of the data set (all N- and/or P- (co-)limited,  $n$  = 539), species richness of vascular plants is highest at intermediate N:P ratios (N and P co-limitation; Fig. 1a), whereas the numbers of endangered species tend to be highest at higher N:P ratios (Fig. 1b) and the percentage of endangered species

gives abbreviations and trait units. Linear regression models are shown, plus their standardized coefficient values ( $\beta$ ) and two-tailed  $P$  values of the coefficients ( $***P < 0.001$ ; NS, not significant). Bar graphs show fraction of variance (for continuous traits) or deviance (for binary traits) of community-mean trait values explained by productivity and N:P ratio. Variances in community-mean trait values are separated into unique effects of productivity (green) and N:P ratio (purple), and shared effects of productivity and N:P ratio (grey). Negative shared effects indicate that the trait is suppressed by interaction between productivity and N:P ratio. For these traits, total variance explained (dotted lines) is smaller than the sum of unique effects of N:P ratio and productivity.

risers as N:P ratios increase towards P limitation (Fig. 1c). For 446 herbaceous plant species occurring in the 539 plots, we retrieved 15 life-history plant traits related to nutrient acquisition, growth and reproductive strategy (see Extended Data Table 1). After accounting for potentially confounding effects of productivity on N:P ratio and community-mean trait values (see Extended Data Fig. 1), we found that plots with a high N:P ratio were significantly ( $P < 0.001$ ) associated with the occurrence of species with a small investment in sexual reproduction (such as low seed number (Fig. 2f) and seed investment (Fig. 2g), late start of flowering (Fig. 2h), short flowering period (Fig. 2i), vegetative reproduction rather than reproduction by seeds (Fig. 2k, l), perennials rather than

annuals or biennials (Fig. 2m)) and less association with N fixers (Fig. 2o). Species occurring at high N:P ratio plots also had leaf traits characteristic of slow-growing species, such as low specific leaf area (SLA) (Fig. 2c) and high leaf dry-matter content (LDMC) (Fig. 2d). Some traits (such as canopy height, leaf mass and plant architecture (monocot or eudicot)) correlated strongly ( $P < 0.001$ ) with productivity but not ( $P > 0.05$ ) with N:P ratio, as shown by a dominant contribution of productivity to their explained variance (Fig. 2a, b, n). The relationships between N:P ratio and particular plant traits also became apparent when species were classified into 'strategies' according to a previous paper<sup>18</sup>; that is, the N:P ratios correlated positively with abundance of 'stress tolerators' ( $P < 0.001$ ), correlated negatively with abundance of 'ruderal' ( $P < 0.001$ ), and were not significantly correlated with abundance of 'competitors' ( $P > 0.05$ ) (Fig. 3). Furthermore, a principal component analysis (PCA) indicated that the trade-off between seed and vegetative reproduction was strongly correlated ( $P < 0.001$ ) with N:P ratio, with the unique effect of N:P ratio being stronger (21% of variation explained) than that of productivity (9% of variation explained), a finding that confirmed the robustness of our analysis (See Supplementary Discussion 2).

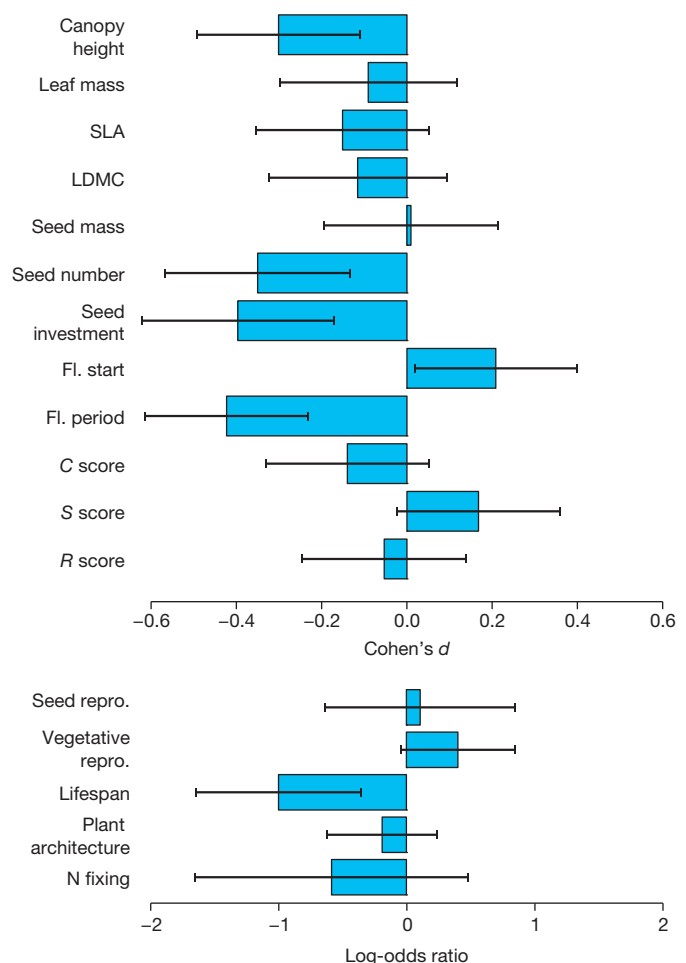
We found three contrasting but not mutually exclusive relationships between nutrient availability or stoichiometry and life-history plant traits. First, overall nutrient availability—but not N:P stoichiometry—is related to traits involved in competition for light (for example, plant size, 'competitor' strategy). Second, both N:P stoichiometry and overall nutrient availability are related to leaf economy traits (for example, specific leaf area, leaf dry-matter content). Previous studies have shown that leaf economy traits are related to overall nutrient availability<sup>17,19</sup>, but our results show that fast-growing species also have an affinity for low N:P ratio environments, independently of the overall nutrient availability effect. Third, N:P stoichiometry is related to investment in sexual reproduction almost independently of overall nutrient availability; that



**Figure 3 | Relationship between N:P ratios and community-mean values for C, S and R scores of CSR strategy after removing confounding effects of productivity.** a, C (competitor) scores. b, S (stress tolerator) scores. c, R (ruderal) scores. Confounding effects of productivity were removed using the residual values of N:P ratio (x axis) and community-mean CSR scores (y axis) regressed by productivity. Number of sites ( $n$ ) = 528; see Fig. 2 legend for figure specifications.

is, high N:P conditions (P limitation) correlated with low investment in sexual reproduction. Such low investment restricts P losses, since reproductive organs are P-rich<sup>20</sup>. Classical allocation studies have shown that plants can invest up to 50 to 60% of all acquired P in sexual reproduction, and that this percentage is generally higher for P than for N and other elements<sup>21,22</sup>. Impaired investment in sexual reproduction under P limitation is also supported by experimental data<sup>23</sup>. An alternative strategy for a plant is to produce few seeds in order to maintain a high P concentration per seed, which is an important factor for successful recruitment in P-impooverished soils<sup>24</sup>.

Our study clearly shows that endangered species have different suites of functional traits than non-endangered species (permutational multivariate analysis of variance (PERMANOVA),  $F_{1,281} = 2.67$ ,  $P < 0.05$ ). Compared with non-endangered species, they have a lower canopy height, less investment in sexual reproduction (fewer seeds and smaller seed mass per individual), a shorter flowering period and a later starting time of flowering, and are perennials rather than annuals (all differences significant at least at the  $P < 0.05$  level) (Fig. 4). The lower reliance of endangered species on seed reproduction is also seen in the PCA axis scores (Supplementary Discussion 2), and confirms the findings of previous



**Figure 4 | Difference in trait values between endangered and non-endangered herbaceous vascular plant species.** The difference was expressed with Cohen's  $d$  for continuous traits, and with log-odds ratios for binary traits. The number of endangered versus non-endangered species examined are (from top to bottom); 150/267, 123/237, 128/246, 121/239, 126/243, 113/226, 101/223, 153/265, 153/265, 149/259, 149/259, 149/259, 149/253, 149/253, 155/268, 152/281 and 157/289. Positive values of Cohen's  $d$  or log-odds ratios indicate that endangered species have higher trait values. 95% confidence intervals are shown as whiskers. Lateral spread, for which effect size could not be calculated, was not significantly different between endangered and non-endangered species (two-tailed Mann-Whitney U-test,  $P > 0.05$ ).

studies comparing common with endangered or rare species (for example, a shorter flowering period<sup>25</sup>, smaller seed mass<sup>26</sup> and poorer dispersal ability<sup>27,28</sup>).

Endangered species occur more frequently under P-limited conditions (high N:P ratio environments) than can be explained by chance, as shown for temperate regions in our data set and in a previously studied much more limited data set<sup>2</sup> as well as for a tropical region<sup>8</sup>. Our trait analysis provides two possible explanations for the frequent occurrence of endangered species in P-limited conditions. The first is that endangered species are often small and are therefore poor competitors for light. Small size is a major disadvantage when growing in productive sites, but on poorly productive sites, which are associated with a high plant N:P ratio (see Supplementary Discussion 1), they face little competition. Second, the relatively low investment of endangered species in sexual reproduction is characteristic of plant species under P limitation (high N:P ratio environments). Thus, both increased productivity of ecosystems and changed N:P stoichiometry potentially threaten the survival of such species; moreover, their low dispersal capacity makes them vulnerable to such threats. The idea that endangered species are vulnerable for changes in the relative availabilities of N and P is supported by a global study showing that species with a narrow geographical range (that is, those more likely to become endangered) have higher leaf N:P ratios than those with a wide range<sup>29</sup>. The exact mechanisms and potentially interacting processes that may explain why species vulnerable for extinction occur on P-limited sites need to be tested further. However, it is likely that large-scale P enrichment of herbaceous ecosystems that boosts productivity and ends P limitation causes species adapted to P limitation to be more vulnerable to extinction. Moreover, the low investment in sexual reproduction of these species, which is a beneficial trait in P-poor environments, is a drawback for their dispersal ability. N fertilization will probably not promote survival of endangered species, as there are a number of mechanisms for increasing P uptake from diverse forms of P in soil (for example, root exudates, mycorrhizae) under N-rich conditions<sup>30</sup>, and therefore P limitation might not be enhanced by N enrichment. Instead, to better protect endangered species, we should aim to preserve P-limited and poorly productive sites. Given that these sites are already scarce and scattered, that landscapes are increasingly human-influenced and urbanized, and that endangered species have less sexual reproduction (and so are disadvantaged in long-distance dispersal), it is clear that these species' vulnerability for extinction is acute.

## METHODS SUMMARY

Species composition of vascular plants, their aboveground biomass (as a proxy for site productivity), and N:P:K ratio in the biomass (as a proxy for relative availability of N, P and K for plants) were recorded in 599 plots in herbaceous ecosystems. The selected ecosystems range from wet to moist conditions and include grasslands, fens, bogs, marshes, reedland, floodplains and dune slacks in nine countries in Eurasia. As N:P stoichiometry is our focus, K-limited plots (N:K ratio >2.1 and K:P ratio <3.41;  $n = 60$ ) were excluded from further analysis. Of the total 491 vascular plant species recorded, 172 endangered species were identified from the combined Red Lists of seven of the countries. We examined the effects of N:P ratio on biodiversity indices (number of species, number and percentage of endangered species) and community-mean values of 15 functional traits and Grime's CSR (competitor, stress tolerator, ruderal) strategy of 446 herbaceous species retrieved from trait databases (see Extended Data Table 1). In all analyses, the confounding effects of productivity on the variables of interest were statistically removed (Extended Data Fig. 1). For the effects of N:P ratio on biodiversity indices, quantile regression analysis was carried out between N:P ratios (corrected for productivity) and biodiversity indices for the 0.50 to 0.95 quantiles. The impact of N:P ratio on community-mean trait values (both corrected for productivity) was tested using path analysis, and the relative contributions of productivity and N:P ratio to community-mean trait values were quantified by partitioning the explained variance of traits to unique and shared effects of productivity and N:P ratio. Furthermore, differences between endangered and non-endangered species in terms of their functional trait composition were examined with a PERMANOVA. In addition, for each of 15 traits and CSR strategy, the differences between endangered and non-endangered species were examined by Cohen's  $d$  for continuous traits and log-odds ratio for binary traits.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 July; accepted 3 October 2013.

Published online 17 November 2013.

1. Olde Venterink, H., Wassen, M. J., Verkroost, A. W. M. & de Ruiter, P. C. Species richness–productivity patterns differ between N-, P-, and K- limited wetlands. *Ecology* **84**, 2191–2199 (2003).
2. Wassen, M. J., Olde Venterink, H., Lapshina, E. D. & Tanneberger, F. Endangered plants persist under phosphorus limitation. *Nature* **437**, 547–550 (2005).
3. Ceulemans, T., Merckx, R., Hens, M. & Honnay, O. A trait-based analysis of the role of phosphorus vs. nitrogen enrichment in plant species loss across North-west European grasslands. *J. Appl. Ecol.* **48**, 1155–1163 (2011).
4. Klaus, V. H. *et al.* Nutrient concentrations and fibre contents of plant community biomass reflect species richness patterns along a broad range of land-use intensities among agricultural grasslands. *Persp. Plant Ecol. Evol. System.* **13**, 287–295 (2011).
5. Cardinale, B. J., Hillebrand, H., Harpole, W. S., Gross, K. & Ptacnik, R. Separating the influence of resource 'availability' from resource 'imbalance' on productivity–diversity relationships. *Ecol. Lett.* **12**, 475–487 (2009).
6. Elser, J. J. *et al.* Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecol. Lett.* **10**, 1135–1142 (2007).
7. Olde Venterink, H., Wassen, M. J., Belgers, D. M. & Verhoeven, J. T. A. Control of environmental variables on species density in fens and meadows: importance of direct effects and effects through community biomass. *J. Ecol.* **89**, 1033–1040 (2001).
8. Lannes, L. S., Bustamante, M. M. C., Edwards, P. J. & Olde Venterink, H. Alien and endangered plants in the Brazilian Cerrado exhibit contrasting relationships with vegetation biomass and N:P stoichiometry. *New Phytol.* **196**, 816–823 (2012).
9. Lambers, H. & Poorter, H. Inherent variation in growth rate between higher plants: a search for physiological causes and ecological consequences. *Adv. Ecol. Res.* **23**, 187–261 (1992).
10. Aerts, R. & Chapin III, F. S. The mineral nutrition of wild plants revisited: a re-evaluation of processes and patterns. *Adv. Ecol. Res.* **30**, 1–67 (2000).
11. Güsewell, S. Nutrient resorption of wetland graminoids is related to the type of nutrient limitation. *Funct. Ecol.* **19**, 344–354 (2005).
12. Güsewell, S. High nitrogen: phosphorus ratios reduce nutrient retention and second-year growth of wetland sedges. *New Phytol.* **166**, 537–550 (2005).
13. Lambers, H., Raven, J. A., Shaver, G. R. & Smith, S. E. Plant nutrient-acquisition strategies change with soil age. *Trends Ecol. Evol.* **23**, 95–103 (2008).
14. Olde Venterink, H. & Güsewell, S. Competitive interactions between two meadow grasses under nitrogen and phosphorus limitation. *Funct. Ecol.* **24**, 877–886 (2010).
15. Fujita, Y., Robroek, B. J. M., de Ruiter, P. C., Heil, G. W. & Wassen, M. J. Increased N affects P uptake of eight grassland species: the role of root surface phosphatase activity. *Oikos* **119**, 1665–1673 (2010).
16. Olde Venterink, H. Does phosphorus limitation promote species-rich plant communities? *Plant Soil* **345**, 1–9 (2011).
17. Elser, J. J. *et al.* Biological stoichiometry from genes to ecosystems. *Ecol. Lett.* **3**, 540–550 (2000).
18. Grime, J. P., Hodgson, J. G. & Hunt, R. *Comparative Plant Ecology: a Functional Approach to Common British Species*. (Castlepoint Press, 2007).
19. Ordóñez, J. C. *et al.* A global study of relationships between leaf traits, climate and soil measures of nutrient fertility. *Glob. Ecol. Biogeogr.* **18**, 137–149 (2009).
20. Kerkhoff, A. J., Fagan, W. F., Elser, J. J. & Enquist, B. J. Phylogenetic and growth form variation in the scaling of nitrogen and phosphorus in the seed plants. *Am. Nat.* **168**, E103–E122 (2006).
21. Fenner, M. The allocation of minerals to seeds in *Senecio vulgaris* plants subjected to nutrient shortage. *J. Ecol.* **74**, 385–392 (1986).
22. Van Andel, J. & Vera, F. Reproductive allocation in *Senecio Sylvaticus* and *Chamaenerion Angustifolium* in relation to mineral nutrition. *J. Ecol.* **65**, 747–758 (1977).
23. Vergeer, P., Rengelink, R., Joop Ouborg, N. & Roelofs, J. G. M. Effects of population size and genetic variation on the response of *Succisa pratensis* to eutrophication and acidification. *J. Ecol.* **91**, 600–609 (2003).
24. Lambers, H., Brundrett, M. C., Raven, J. A. & Hopper, S. D. Plant mineral nutrition in ancient landscapes: high plant species diversity on infertile soils is linked to functional diversity for nutritional strategies. *Plant Soil* **334**, 11–31 (2010).
25. Lahti, T., Kemppainen, E., Kurtto, A. & Uotila, P. Distribution and biological characteristics of threatened vascular plants in Finland. *Biol. Conserv.* **55**, 299–314 (1991).
26. Murray, B. R., Thrall, P. H., Gill, A. M. & Nicotra, A. B. How plant life-history and ecological traits relate to species rarity and commonness at varying spatial scales. *Austral Ecol.* **27**, 291–310 (2002).
27. Kunin, W. E. & Gaston, K. J. The biology of rarity—patterns, causes and consequences. *Trends Ecol. Evol.* **8**, 298–301 (1993).
28. Farnsworth, E. J. & Oğurcak, D. E. Functional groups of rare plants differ in levels of imperilment. *Am. J. Bot.* **95**, 943–953 (2008).

29. Elser, J. J., Fagan, W. F., Kerkhoff, A. J., Swenson, N. G. & Enquist, B. J. Biological stoichiometry of plant production: metabolism, scaling and ecological response to global change. *New Phytol.* **186**, 593–608 (2010).
30. Vance, C. P., Uhde-Stone, C. & Allan, D. L. Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. *New Phytol.* **157**, 423–447 (2003).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We would like to thank N. A. Soudzilovskaia for obtaining the Russian Red List of plant species, M. Soons, D. Ertzen and D. van der Goes for permission to use their vegetation records, T. Markus and M. Stoete for drawing figures and J. Burrough for editing the near-final draft. Y.F. was funded by the Utrecht Centre of

Geosciences, and the research in the Rospuda river valley was financed by the Polish Ministry of Science and Higher Education Grant no. N304 010 31/0414.

**Author Contributions** Y.F., H.O.V., N.H., E.J., W.K., P.P., T.O. and M.J.W. collected data; Y.F., H.O.V., P.M.v.B., J.C.D. and M.J.W. analysed data; Y.F., H.O.V., P.M.v.B., P.C.d.R. and M.J.W. wrote the manuscript; J.C.D., G.W.H., N.H., E.J., W.K., P.P. and T.O. commented on the manuscript; P.C.d.R., G.W.H. and M.J.W. were project leaders.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.W. ([m.j.wassen@uu.nl](mailto:m.j.wassen@uu.nl)).



## METHODS

**Plot selection.** We collected vegetation data on 599 plots from herbaceous ecosystems, including the 276 sites studied in ref. 2. The sample consists of grasslands, fens, bogs, marshes, reed beds and dune-slack vegetation. All plots are non-brackish, with moist to wet conditions, thus impacts of drought and salinity on species richness were avoided. Only 19 plots are lightly fertilized ( $\leq 100 \text{ kg N ha}^{-1} \text{ yr}^{-1}$ ), and 102 plots are exposed to periodic river flooding. All plots are dominated by herbaceous species (>50% cover). The plots were selected to span a wide geographical range of Eurasian countries: The Netherlands (255 plots), Poland (153), Russia (82), Germany (43), Belgium (20), Iceland (17), Sweden (10), Scotland (10) and Belarus (9). These plots encompass most of the west and central European lowlands. The flora in the country-wise partial data sets we collected is similar: on average, 86% (and a minimum of 70%) of the species recorded in any country had also been recorded in the other countries. For each plot, aboveground standing biomass of vascular plants was harvested at the peak of the growing season; that is, from June to August. The harvested area ranged from  $0.06 \text{ m}^2$  to  $1 \text{ m}^2$ . The content of N, P and K in the biomass was analysed after Kjeldahl digestion<sup>31</sup>. Composition of vascular plant species was recorded in or around the harvested area in plots of  $0.06$  to  $25 \text{ m}^2$ . The different plot sizes did not affect the relationships between biodiversity and N:P stoichiometry (see Supplementary Discussion 3).

**Functional characteristics of species.** Per species, we quantified those functional traits available in trait databases<sup>32–38</sup> for most of the recorded plant species that are known to be important for the growth or reproduction strategies of herbaceous plant species<sup>39</sup> (See Extended Data Table 1 for an overview of traits, units and sources). We excluded woody species (45 species out of the total 491 species) from the trait analysis, as most woody species recorded in our plots were seedlings and therefore the trait values available in databases (which are for adult individuals) are not relevant. The selected traits were those related to competition for light (canopy height, leaf mass), leaf economy traits (specific leaf area (SLA), leaf dry-matter content (LDMC)), seed traits (seed mass, number of seeds per individual, seed investment (that is, seed mass per individual, calculated as seed mass times number of seeds per individual)), phenology traits (starting month of flowering, duration of flowering period), reproduction strategy traits (lateral spread, type of reproduction, plant lifespan), a plant architecture trait (that is, eudicots or monocots), and a nutrient acquisition trait (N fixation) (Extended Data Table 1). We note that seed investment is a trait that may be biased by the size of the plant. However, even when we corrected seed investment roughly for plant size, the relationship between N:P ratio and this trait did not change (Supplementary Discussion 4). Duration of flowering period (expressed in months) and starting month of flowering (ranging from January to August, and thus coded 1 to 8) were treated as continuous variables. All continuous traits except LDMC and starting month of flowering were log-transformed to adjust the right-skewed frequency distributions. Type of reproduction, expressed on an ordinal scale with five classes, was converted into two binary variables (reproduction by seeds and vegetative reproduction) to be used in subsequent analyses. Additionally, the CSR strategy<sup>18</sup> was attributed to each species by using seven traits (canopy height, LDMC, leaf mass, SLA, flowering period, starting month of flowering, and lateral spread) according to a method described previously<sup>40</sup>. When this method could not be applied because trait information was incomplete, the CSR classification in BioFlor<sup>36</sup> was used. Ultimately, CSR classes were defined for 408 out of 446 herbaceous species. For each species, scores for each primary component (C, competitor; S, stress tolerator; R, ruderal) were assigned from its proportional contributions (for example, C scores are 1 for the strategy 'C', 0.5 for 'CS', 0.33 for 'CSR' and 0.75 for 'C/CR'<sup>41</sup>). In addition, we used a principal component analysis (PCA) to identify the major axes of variation in multiple functional traits (see Supplementary Discussion 2).

For each plot, community-mean values (unweighted for the abundance of species) of continuous traits were calculated as an indicator of the mean response of the plant community to the site conditions. We treated the community-mean values of lateral spread, an ordinal trait, in the same way as those of continuous traits, because they were approximately normally distributed. For binary traits, the number of species with 1s and 0s were counted per plot. Plots with fewer than three species with a valid trait value, and plots in which less than 50% of occurring herbaceous species had a valid trait value, were omitted from the analysis. The omitted plots ranged from 6 to 16 plots (average 10.7 plots) of the total (539 plots).

**Endangered species.** We compiled a list of endangered species by combining the regional Red Lists of the Netherlands<sup>42</sup>, Germany<sup>43</sup>, Poland<sup>44</sup>, Sweden<sup>45</sup>, UK<sup>46</sup>, Iceland<sup>47</sup> and the Novosibirsk region in Siberia<sup>48</sup>. The Red List of Belgium was not included, because the Belgian plots were near the border with the Netherlands and their flora was comparable with the flora in the Dutch plots. We also did not use the list of Belarus, because the number of plots in this country was small ( $n = 9$ ) and the flora in these plots overlapped with those of Poland. The Red List status of a species reflects both the decline of the habitat in the region and the susceptibility of the species to the changing environment. Some species are Red

Listed in one country but not in others, because in those countries their habitat is not deteriorating. For our analysis, only the susceptibility of an endangered species is relevant, not the region-specific habitat deterioration, because we are interested in the mechanisms whereby species become endangered (that is, their functional traits). Therefore, we identified a species as 'endangered' if it is on at least one regional Red List (meaning that this species has fragile characteristics that are susceptible to environmental changes), and applied this new list to all plots. In this way we corrected for habitat loss, which is largely region-specific. Note that we included the categories which refer to actual decline of the species ('critically endangered', 'endangered' and 'vulnerable' species), but not the category which refers to the scarcity of the species ('rare'). In this way, we excluded species which are always rare irrespective of environmental change. Of our 491 vascular plant species, we identified 172 endangered species (157 herbaceous and 15 non-herbaceous species). For each plot, we counted number and percentage of endangered species. The latter is the number of endangered species divided by the total number of species ( $\times 100$ ) per plot.

**Data analysis approach.** We are interested in the effects of nutrient stoichiometry (that is, the ratio between N and P availability for plants) on species diversity and on prevailing functional traits of species in plant communities, irrespective of confounding effects of overall nutrient availability. We use N:P ratios of aboveground plant biomass as a proxy for nutrient stoichiometry (see Supplementary Discussion 1 for a justification of using plant N:P ratio), and site productivity (that is, aboveground biomass of vascular plants) as a proxy for overall nutrient availability. As the consideration of N:P ratio is relevant only when a plot is limited or co-limited by N or P, we excluded K-limited plots ( $n = 60$ ) from the analysis. We considered a plot as K-limited if the N:K ratio was more than 2.1 and the K:P ratio was less than 3.4 (ref. 1).

We propose the following relationships between plant N:P ratio, site productivity, species diversity and community-mean species traits (see also Extended Data Fig. 1). Plant N:P ratio and site productivity are related through a scaling law proposed by the growth rate hypothesis<sup>29,49</sup> (arrow *a* in Extended Data Fig. 1): plants grown in fertile environments (these tend to be fast-growing species) exhibit low biomass N:P ratios because of the high amount of P-rich RNA needed for rapid division of cells<sup>50,51</sup> (see Supplementary Discussion 1 for more details). We are aware of the opposite direction of effect too (that is, N:P ratio influencing site productivity), particularly at extreme values of N:P ratios, where a deficiency of N or P limits site productivity, but we consider this effect to be minor. Furthermore, we posit that both species diversity and community-mean trait values are influenced by site productivity and N:P ratio. However, there is an intrinsic difference between how these drivers affect species richness and how they affect community-mean traits. Site productivity and N:P ratio 'filter' the community-mean traits (arrows *e* and *d*, respectively, in Extended Data Fig. 1b), acting on the mean of the community-mean trait values. In contrast, site productivity and N:P ratio 'limit' species diversity (arrows *c* and *b*, respectively, in Extended Data Fig. 1a), acting on the upper values of species diversity. Given the different nature of the relationships, we employed two sets of methods to eliminate potential confounding factors from our analysis. All analyses were performed in R<sup>52</sup>.

**Effects of N:P ratio on species diversity.** A relationship between a dependent variable (species diversity) and one or more limiting factors (site productivity and N:P ratio) can be tested with quantile regression analysis. Unlike conventional regression, which considers solely changes in the mean of the response variable, quantile regression excludes the effect of unmeasured limiting factors<sup>53</sup>. It is therefore a powerful method to analyse the change in the potential species diversity as a function of the limiting factor; that is, N:P ratio, only.

We used three indices to assess species diversity: number of species ( $y_1$ ), number of endangered species ( $y_2$ ), and percentage of endangered species over total number of species ( $y_3$ , which equals to  $100 \cdot y_2 / y_1$ ). For the number of species and number of endangered species, we assumed a quadratic effect of N:P ratio, since previous studies suggested that on an N:P ratio gradient there is an optimum biodiversity rather than a continuously increasing biodiversity<sup>2</sup>. Both response variables were log-transformed (that is,  $\ln(y_1)$  and  $\ln(y_2 + 1)$ ) to correct the right-skewed frequency distributions. For the percentage of endangered species, a logistic quantile regression analysis was applied to restrict the prediction to between 0 and 100. Here we used 'empirical logits'<sup>54</sup>,  $\ln((y_2 + 0.5)/(y_1 - y_2 + 0.5))$ , instead of normal logits,  $\ln(y_2/(y_1 - y_2))$ , to enable computation of logits when a plot has 0% or 100% of endangered species. Effects of N:P ratio on species diversity indices (arrow *b* in Extended Data Fig. 1a), irrespective of the confounding effects of site productivity on N:P ratio (arrow *a* in Extended Data Fig. 1a), were examined by using 'the residual values of N:P ratio versus productivity' (obtained from a linear regression model on a log-log scale) as an explanatory variable.

For each diversity index, the  $\tau$ th quantile regression function  $B_{(\tau)}$  was determined, indicating the approximately  $\tau$ th proportion of the observations found to be below  $B_{(\tau)}$  (ref. 55). As we are interested in the upper boundary of the relationship

when N:P ratio actively limits biodiversity, we examined high values of  $\tau$  only ( $\tau > 0.50$ ). To evaluate the precision of the obtained model, 95% confidence intervals of the coefficients were computed with the rank inversion method<sup>55</sup> (Extended Data Fig. 2). These analyses were performed with the R package 'quantreg'<sup>55</sup>.

The analyses were run for the complete data set. A control run was carried out on the data set excluding two dominant habitat types (fens and bogs). This tested whether our analysis results could have been biased by habitats that may be severely deteriorated (for example, fens and bogs in western Europe with desiccation problems) or that have intrinsically high or low N:P ratios (for example, bogs tend to have high N:P ratios because this rainwater-fed ecosystem receives relatively high N supply through atmospheric deposition but poor P supply) (Extended Data Fig. 3). We did not find evidence for such bias.

We also tested whether other factors potentially related to P limitation (for example, soil acidity, soil moisture) confound the association between N:P ratio and richness of endangered species (Supplementary Discussion 5), but this was not the case.

**Effects of N:P ratio on community-mean traits.** Our aim was twofold: to examine, first, the significance of N:P ratios in affecting community-mean trait values; and second, the explained variation in community-mean trait values by N:P ratios in addition to and in interaction with productivity.

First, we tested the association between N:P ratio and community-mean traits in the proposed relationships between N:P ratio, productivity and community-mean traits (Extended Data Fig. 1b) using concepts of path modelling<sup>56</sup>. Note that we proposed the causal model as depicted in Extended Data Fig. 1b to be true, even though causality as such cannot be tested in this model because the hypothesized model is 'saturated' (that is, all the possible interconnections are specified<sup>57</sup>). However, it is still possible to refute the existence of a relationship between N:P ratio and community-mean trait values (arrow  $d$  in Extended Data Fig. 1b) by testing the significance of this association while statistically holding productivity constant<sup>56</sup>. More concretely, the association between N:P ratio and community-mean values of a trait was examined from the relationship between the residuals of community-mean trait values regressed by productivity ( $\text{res}_{\text{tr},p}$ ) and the residuals of N:P ratio regressed by productivity ( $\text{res}_{\text{NP},p}$ ). When the regression coefficient of  $\text{res}_{\text{tr},p}$  regressed by  $\text{res}_{\text{NP},p}$  was not significantly different from zero, we concluded that there was no significant association between N:P ratio and the community-mean trait.  $\text{res}_{\text{tr},p}$  were derived from logistic models with negative binomial distribution for the binary traits, and from linear models for continuous traits, using the generalized linear model (GLM) (Extended Data Fig. 4). To test the relationship between  $\text{res}_{\text{tr},p}$  and  $\text{res}_{\text{NP},p}$  we assumed linear regression models (after square-root-transformation of  $\text{res}_{\text{tr},p}$  for lifespan and log-transformation of  $\text{res}_{\text{tr},p}$  for N fixation). The residuals of the linear regression models did not deviate from a normal distribution ( $P > 0.05$  with Kolmogorov–Smirnov test except for SLA ( $P = 0.004$ ), lateral spread ( $P = 0.046$ ) and N fixation ( $P = 0.01$ ); for these three traits we also checked the  $\text{res}_{\text{tr},p} - \text{res}_{\text{NP},p}$  relationship with Spearman's correlation analysis: our conclusions were identical). To visualize the effects of N:P ratio on community-mean traits (Fig. 2 and Fig. 3), we have shown  $\text{res}_{\text{tr},p}$  against  $\text{res}_{\text{NP},p}$  and the significance and strength of this relationship ( $\beta$ , the standardized regression coefficient). In a path analysis context,  $\beta$  corresponds to a relationship between N:P ratio and community-mean traits for which direction of the causality is not prescribed. The significance levels of  $\beta$ s and the path coefficients (which prescribe the direction of causality) were nearly identical (results not shown), showing that our conclusion on the existence of association between community-mean traits and N:P ratio is robust against the assumed direction of the causality.

Second, to examine the explained variation in community-mean trait values by N:P ratios in addition to and in interaction with productivity, we partitioned the variation in community-mean trait values into values explained by unique effects of N:P ratio and productivity and their shared effects. We did so by sequentially calculating the measures of goodness-of-fit ( $G$ ) of regression models<sup>58</sup>. In this way, the relative contribution of N:P ratio and productivity on community-mean trait values can be quantified irrespective of the underlying multicollinearity among three variables<sup>59</sup>. As a measure of  $G$ , we used log likelihood for logistic models and  $R^2$  for linear models. Unique effects of N:P ratio were calculated as an increase in  $G$  in the model of community-mean traits regressed by productivity and N:P ratio ( $G(X_1X_2)$ ) compared to that regressed by productivity only ( $G(X_2)$ ). Unique effects of productivity were calculated identically, as an increase in  $G(X_1X_2)$  compared to  $G(X_1)$ . Shared effects of productivity and N:P ratio were calculated as:  $G(X_1) + G(X_2) - G(X_1X_2)$ .

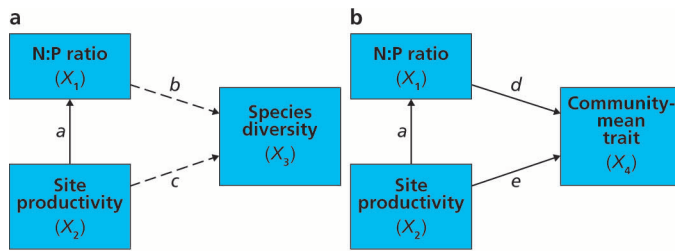
For some traits, shared effects were negative, indicating that the trait is suppressed by interaction between productivity and N:P ratio.

**Traits of endangered species versus non-endangered species.** The difference in a suite of functional traits (11 traits including all the continuous- and ordinal-scale traits) between endangered and non-endangered species was tested with PERMANOVA<sup>60</sup> using the R package 'vegan'<sup>61</sup> for 283 herbaceous species (for which full data were

available for the 11 traits). The distance matrices were calculated based on Gower's distance. The relative difference in each functional trait between endangered and non-endangered species was then quantified by calculating the effect size. Positive values for the effect size mean that endangered species have higher trait values than non-endangered species. For continuous traits, we used Cohen's  $d$  as effect size measure. Canopy height, leaf mass, SLA, seed mass, number of seeds, seed investment, and flowering period were log-transformed before the analysis. The equations of Cohen's  $d$  and its 95% confidence intervals<sup>62</sup> are shown in Supplementary equation (1). For traits that were only marginally normally distributed (flowering start, flowering period, C, S, R scores;  $P < 0.05$  with Kolmogorov–Smirnov test), the difference between endangered and non-endangered species was also tested with the Mann–Whitney U-test; the conclusions were identical. For binary traits, the effect size was examined by the log-odds ratio. The equations of log-odds ratio and its 95% confidence intervals<sup>64</sup> are shown in Supplementary equation (2). Note that the scales of Cohen's  $d$  and log-odds ratio are different, so the effect size of continuous and binary traits cannot be compared. For ordinary traits, effect size cannot be calculated. For these traits, the Mann–Whitney U-test was used to test the difference between endangered and non-endangered species.

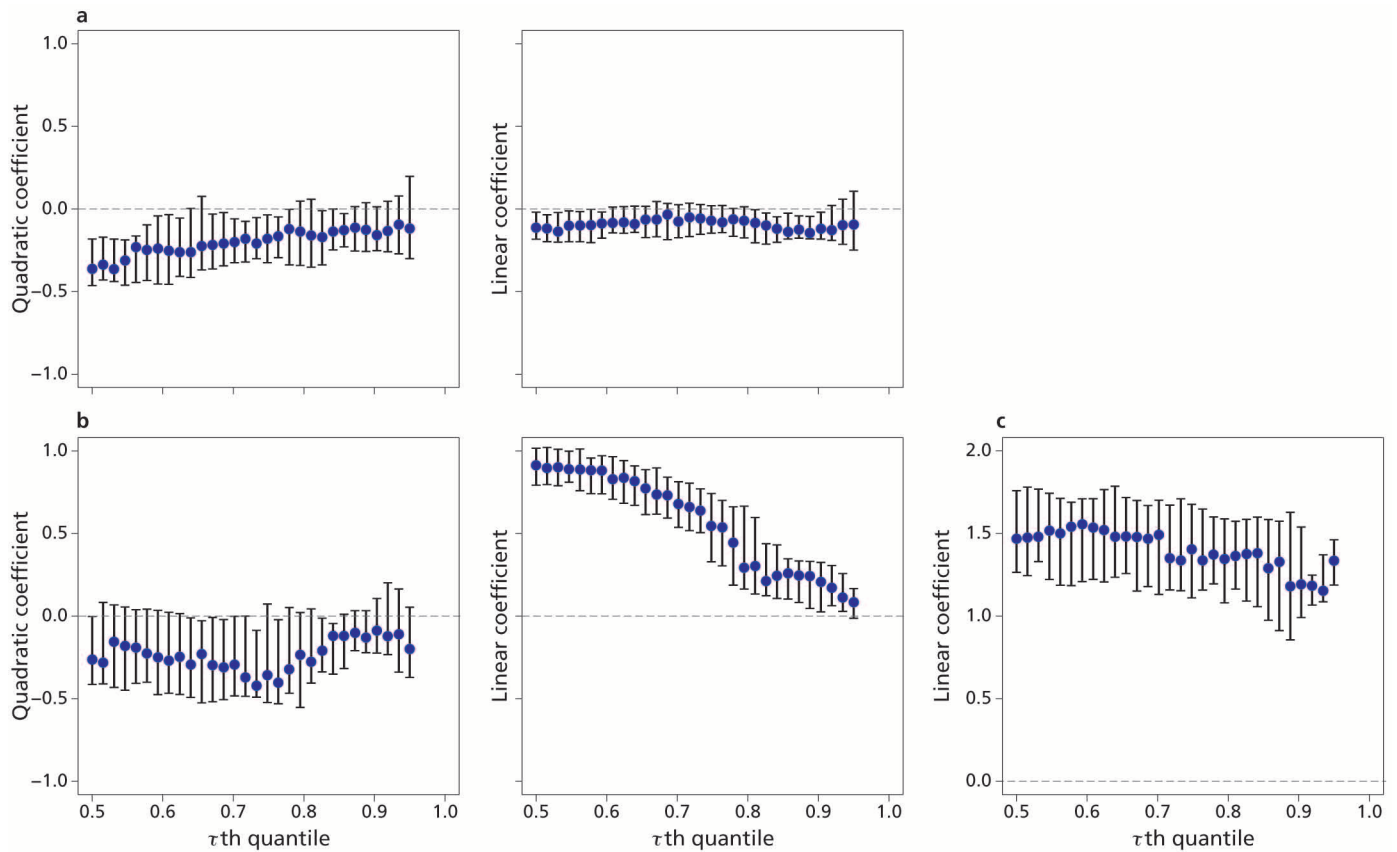
- Bremner, J. M., Mulvaney, C. S., Page, A. C., Miller, R. H. & Keeney, D. R. In *Methods of Soil Analysis; Part 2, Chemical and Microbiological Properties* 621–622 (American society of Agronomy, 1982).
- Centraal Bureau voor Statistiek. *Dutch Botanical Database, BioBase 2003* <http://www.cbs.nl/nl-nl/menu/methoden/classificaties/overzicht/namenlijsten-planten-dieren/2003-biobase.htm> (accessed 18 May 2009).
- Kleyer, M. et al. The LEDA Traitbase: a database of life-history traits of the Northwest European flora. *J. Ecol.* **96**, 1266–1274 (2008).
- Klimešová, J. & de Bello, F. CLO-PLA: the database of clonal and bud bank traits of Central European flora. *J. Veg. Sci.* **20**, 511–516 (2009).
- Hodgson, J. G., Grime, J. P., Hunt, R. & Thompson, K. *The Electronic Comparative Plant Ecology*. (Chapman & Hall, 1995).
- Klotz, S., Kühn, I. & Durka, W. In *BIOLFLOR — Eine Datenbank zu biologisch-ökologischen Merkmalen der Gefäßpflanzen in Deutschland. Schriftenreihe für Vegetationskunde* **38** (Bundesamt für Naturschutz, 2002).
- Sprent, J. I. *Nodulation in Legumes*. (Royal Botanic Gardens, 2001).
- Royal Botanic Gardens Kew. *Seed Information Database (SID). Version 7.1* <http://data.kew.org/sid/09/09/2011>.
- Westoby, M., Falster, D. S., Moles, A. T., Vesk, P. A. & Wright, I. J. Plant ecological strategies: some leading dimensions of variation between species. *Annu. Rev. Ecol. Syst.* **33**, 125–159 (2002).
- Hodgson, J. G., Wilson, P. J., Hunt, R., Grime, J. P. & Thompson, K. Allocating C-S-R plant functional types: a soft approach to hard problem. *Oikos* **85**, 282–296 (1999).
- Willby, N. J., Pulford, I. D. & Flowers, T. H. Tissue nutrient signatures predict herbaceous-wetland community responses to nutrient availability. *New Phytol.* **152**, 463–481 (2001).
- Tamis, W. L. M. et al. Standaardlijst van de Nederlandse flora 2003. *Gorteria* **60**, 101–195 (2004).
- Schnittler, M. & Ludwig, G. In *Rote Listen und Florenlisten gefährdeter Pflanzen in Deutschland. Schriftenreihe für Vegetationskunde*, **28**, 709–739 (Bundesamt für Naturschutz, 1996).
- Mirek, Z., Zarzycki, K., Wojewoda, W. & Szeląg, Z. *Red List of Plants and Fungi in Poland* (Bundesamt für Naturschutz, 2006).
- Swedish Species Information Centre. *ArtDatabanken* <http://www.artfakta.se> (accessed 15 July 2011).
- JNCC. *The Vascular Plant Red Data List for Great Britain* <http://jncc.defra.gov.uk> (accessed 15 July 2011).
- Icelandic Institute of Natural History. *Válisti háplantna* <http://www.ni.is/grodur/valisti> (accessed 18 July 2011).
- State Committee for Environmental Protection of the Novosibirsk region. *The Red List of Plants of Novosibirsk Region* <http://edu.nstu.ru/rcc/konkurs/kkrno/spisok.htm> (accessed 30 June 2011).
- Sterner, R. W. & Elser, J. J. *Ecological Stoichiometry: the Biology of Elements from Molecules to the Biosphere*. (Princeton Univ. Press, 2002).
- Ågren, G. I. The C: N: P stoichiometry of autotrophs — theory and observations. *Ecol. Lett.* **7**, 185–191 (2004).
- Niklas, K. J., Owens, T., Reich, P. B. & Cobb, E. D. Nitrogen/phosphorus leaf stoichiometry and the scaling of plant growth. *Ecol. Lett.* **8**, 636–642 (2005).
- R. A language and environment for statistical computing. Version R-3.0.1. <http://www.R-project.org> (16/05/2013).
- Cade, B. S., Terrell, J. W. & Schroeder, R. L. Estimating effects of limiting factors with regression quantiles. *Ecology* **80**, 311–323 (1999).
- Sokal, R. R. & Rohlf, F. J. *Biometry: the Principles and Practice of Statistics in Biological Research* (W. H. Freeman and Company, 1995).
- quantreg: Quantile Regression. R package version 4.76. <http://CRAN.R-project.org/package=quantreg> (02/12/2011).
- Shipley, B. Confirmatory path analysis in a generalized multilevel context. *Ecology* **90**, 363–368 (2009).
- Grace, J. B. *Structural Equation Modeling and Natural Systems* (Cambridge Univ. Press, 2006).
- Borcard, D., Legendre, P. & Drapeau, P. Partialling out the spatial component of ecological variation. *Ecology* **73**, 1045–1055 (1992).

59. Graham, M. H. Confronting multicollinearity in ecological multiple regression. *Ecology* **84**, 2809–2815 (2003).
60. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).
61. vegan: Community Ecology Package. R package version 1.17-9. <http://CRAN.R-project.org/package=vegan> (released on 31/03/2011).
62. Howell, D. C. *Statistical Methods for Psychology* 8th edn (Wadsworth Publishing, 2012).



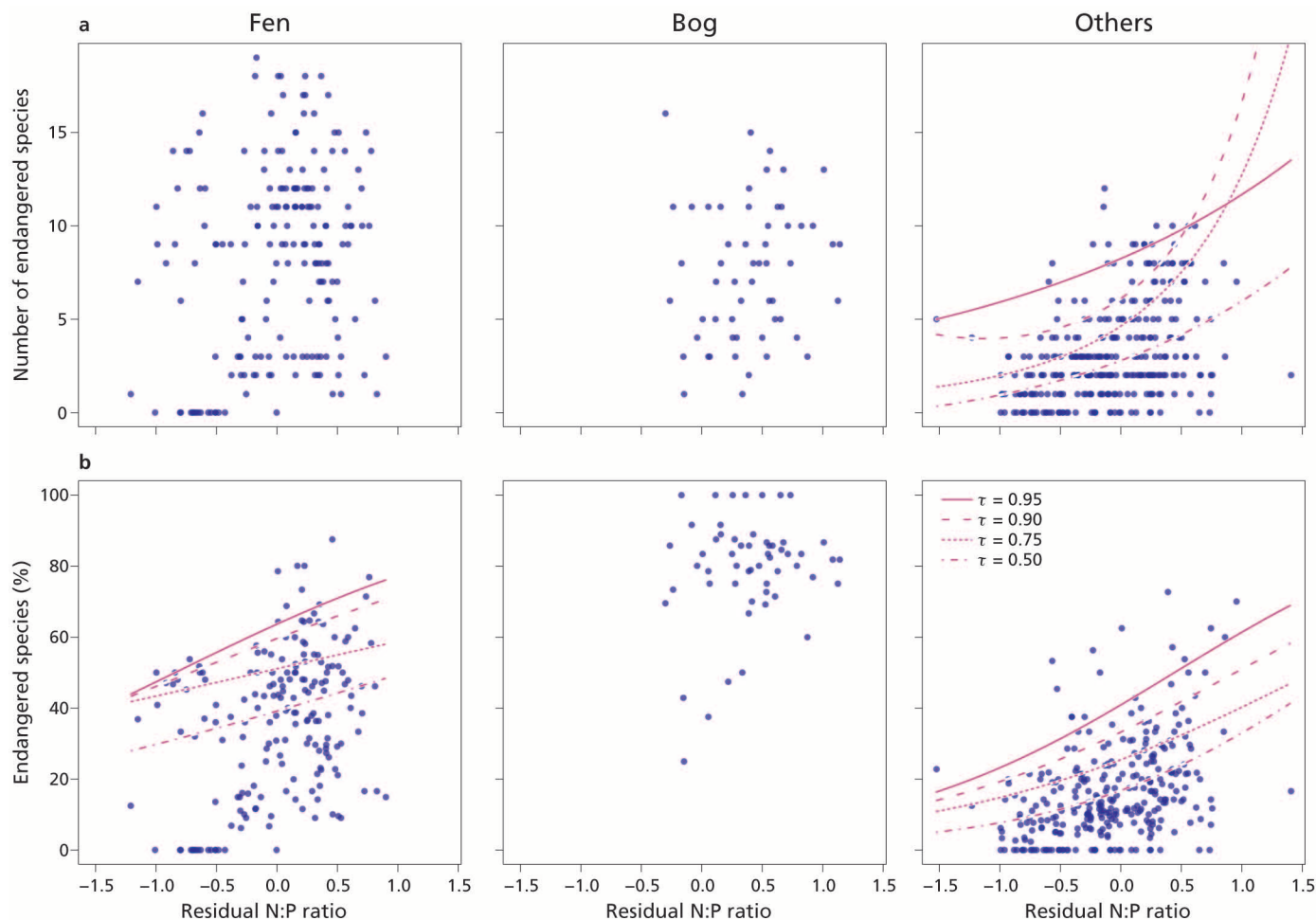
**Extended Data Figure 1 | Data analysis approach.** **a, b,** Schematic proposed relationships between site productivity (that is, aboveground biomass of vascular plants;  $X_2$ ), N:P ratio in aboveground plant biomass ( $X_1$ ), and species diversity (**a**;  $X_3$ ) or community-mean traits (**b**;  $X_4$ ). Solid arrows are relationships in which the explanatory variable is constrained by the response variable (direct causality); dashed arrows are relationships in which upper bound of the explanatory variable is constrained by the response variable (limitation). Arrow *a* represents the pattern predicted by the growth rate hypothesis (see Supplementary Discussion 1 for details). The effect of N:P ratio on species diversity (arrow *b*) was tested by quantile regression analysis (thus treating arrow *c* as another limiting factor) with the residual values of  $X_1$  versus  $X_2$  as an explanatory variable (thus removing the effect illustrated by arrow *a*). The effect of N:P ratio on a community-mean trait (arrow *d*) was tested by comparing the residual values of  $X_1$  versus  $X_2$  (thus removing the effect illustrated by arrow *a*) with the residual values of  $X_4$  versus  $X_2$  (thus removing the effect illustrated by arrow *e*), using concepts of path analysis.





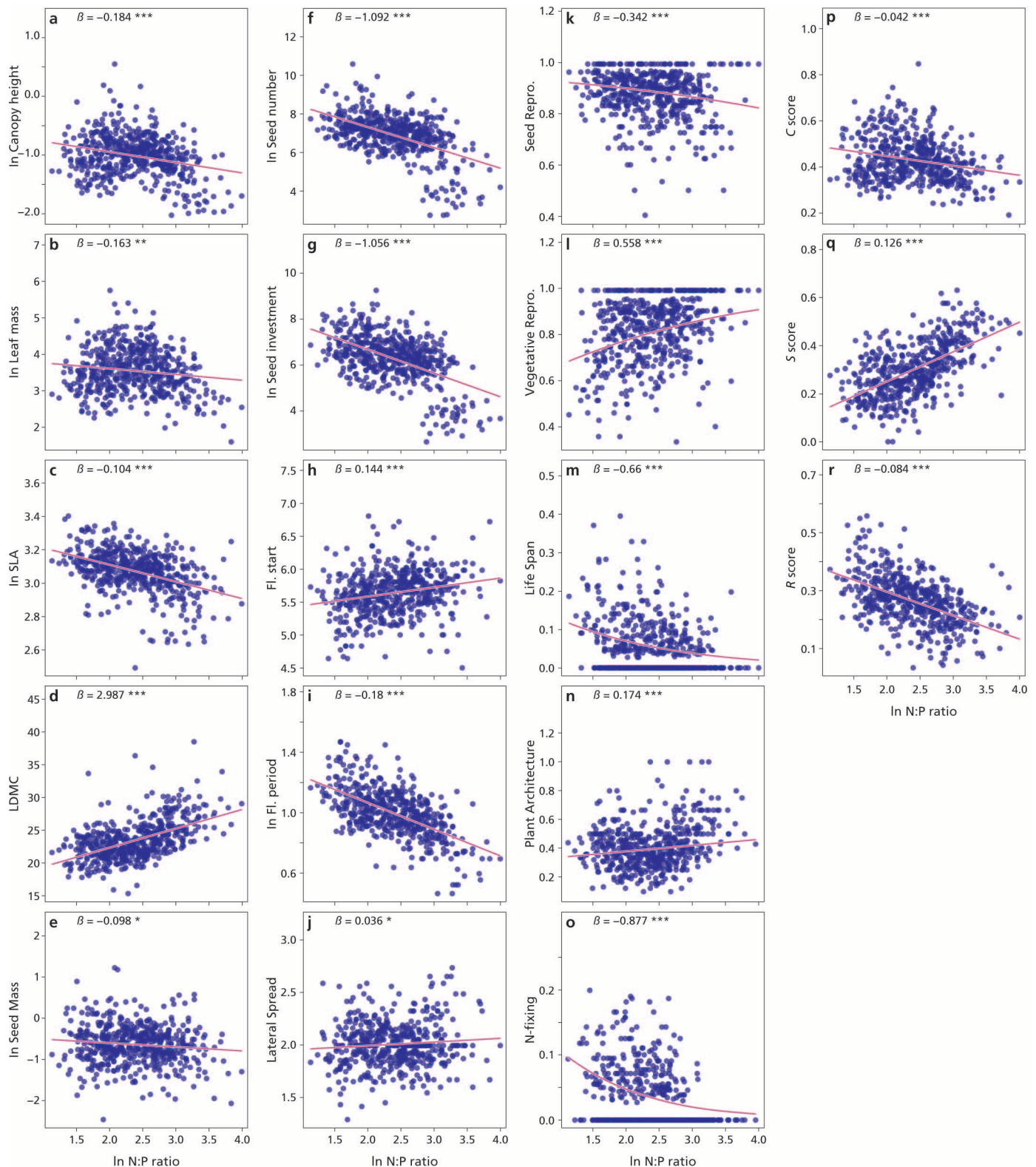
**Extended Data Figure 2 | Ninety-five per cent confidence intervals of the quantile regression coefficients.** a–c, Estimates (dots) and 95% confidence intervals (bars) of quadratic and linear coefficients ( $b_2$  and  $b_1$ , respectively) of quantile regression models are shown for the number of vascular plant species (a), the number of endangered species (b), and the percentage of endangered species (c) regressed by N:P ratio corrected for productivity effects. The fitted models were ( $y_1$ ):  $\ln(y_1) = b_0 + b_1x + b_2x^2$  for number of species;

( $y_2$ ):  $\ln(y_2 + 1) = b_0 + b_1x + b_2x^2$  for number of endangered species; and ( $y_3 = 100 \cdot y_2 / y_1$ ):  $\ln((y_2 + 0.5)/(y_1 - y_2 + 0.5)) = b_0 + b_1x$  for percentage of endangered species, where  $x$  is the residuals of plant N:P ratio regressed by productivity. Models were examined for 50% ( $\tau = 0.50$ ) to 95% ( $\tau = 0.95$ ) quantiles. See Fig. 1 for the shape of the quantile regression models for  $\tau = 0.50, 0.75, 0.90, 0.95$ .



**Extended Data Figure 3 | Effects of habitat types on relationships between residual N:P ratio and biodiversity indices.** Relationships between N:P ratio corrected for productivity effects and the number of endangered species (a) and percentage of endangered species (b) are shown for different habitat types (left, 187 fens; middle, 56 bogs; and right, 296 other habitat types). Linear, rather than quadratic, quantile regression models were applied because for most quantiles the quadratic coefficients did not differ significantly from zero.  $\tau$ th linear quantile regression models ( $\tau = 0.50, 0.75, 0.90, 0.95$ ) are shown only

when the 95% confidence intervals of the linear coefficients of the regression models were above or below zero for the majority of the quantiles. Number and percentage of endangered species increased concomitantly with increasing N:P ratio (corrected for productivity) even in plots that are not fens and bogs, indicating that our findings on the relationship between N:P ratio and endangered species were not an artefact resulting from the stratified sampling of habitat types.



**Extended Data Figure 4 | Relationships between community-mean trait values and plant N:P ratio.** a–r, The tested traits are canopy height (a, number of sites ( $n$ ) = 530), leaf mass (b,  $n$  = 525), specific leaf area (c,  $n$  = 529), leaf dry-matter content (d,  $n$  = 525), seed mass (e,  $n$  = 533), seed number per shoot (f,  $n$  = 524), seed investment (g,  $n$  = 523), starting month of flowering (h,  $n$  = 528), flowering period (i,  $n$  = 528), lateral spread (j,  $n$  = 526), reproduction by seeds (k,  $n$  = 528), vegetative reproduction (l,  $n$  = 528), life span (m,  $n$  = 531), plant architecture (n,  $n$  = 533), N fixation (o,  $n$  = 502), C score (p,  $n$  = 528), S score (q,  $n$  = 528) and R score (r,  $n$  = 528). See Extended

Data Table 1 for abbreviations and units of the traits. Canopy height, leaf mass, specific leaf area, seed mass, number of seeds, seed investment, and flowering period were log-transformed before the calculation of community-mean values. For binary traits, plot mean values were shown as a fraction of species with 1s over total species (that is, sum of 1s and 0s) to allow graphical presentation. Standardized regression coefficients ( $\beta$ ) of community-mean trait regressed by N:P ratio using GLM and their two-tailed  $p$ -values ( $***P < 0.001$ ,  $**P < 0.01$ ,  $*P < 0.05$ ) are shown.

**Extended Data Table 1 | List of functional traits of herbaceous vascular plant species.** 14 species functional traits were retrieved from trait databases, and 2 binary traits ('Reproduction by seeds' and 'Vegetative reproduction') were derived from an ordinal trait ('Type of reproduction'). For regression analyses of community-mean trait values (Fig. 2) and for effect-size calculation (Fig. 4), we did not use 'Type of reproduction' but the two derived binary traits instead. The traits were retrieved for 446 herbaceous vascular plant species. Canopy height, leaf mass, SLA, seed mass, number of seeds, seed investment, and flowering period were log-transformed before all analyses.

Trait	Scale and Unit	% of species with a trait value	Source
Canopy height	Continuous; m	93	32, 33
Leaf mass	Continuous; mg	81	33
Specific leaf area (SLA)	Continuous; mm <sup>2</sup> /mg	84	33
Leaf dry matter content (LDMC)	Continuous; %	81	33
Seed mass	Continuous; mg per seed	83	33, 36, 38
Number of seeds	Continuous; number per shoot	76	33
Seed investment	Continuous; mg per shoot	73	Seed mass x Number of seeds
Starting month of flowering	Continuous*†; month	94	32, 36
Duration of flowering period	Continuous*; month	94	32, 36
Lateral spread	Ordinal‡; 0: annuals, 1: <0.01m, 2: 0.01-0.25m, 3: >0.25m	85	34, 35
Type of reproduction §	Ordinal; 1: s, 0.75: ssv, 0.5: sv, 0.25: svv, 0: v	90	36
Reproduction by seeds	Binary; 1: yes (s/ssv/sv), 0: no or seldom (svv/v)	90	Derived from 'Type of reproduction'
Vegetative reproduction	Binary; 1: yes (v/vvs/sv), 0: no or seldom (ssv/s)	90	Derived from 'Type of reproduction'
Life span	Binary; 1: annual or biennial, 0: Perennial	95	32, 33, 36
Plant architecture	Binary; 1: monocot, 0: eudicot	97	32
N-fixation	Binary; 1: nodulated legume, 0: others	100	37

\* Month is strictly speaking an ordinal scale, but treated here as a continuous scale.

† Range from January to August, thus coded as 1 to 8.

‡ Community-mean values of lateral spread were treated as continuous variables, as lateral spread does not deviate from normal distribution.

§ Expressed as the relative dependency on seed reproduction. The original categories in the database are: s (by seeds), ssv (mostly by seeds), sv (both by seeds and vegetatively), vvs (mostly vegetatively), v (vegetatively).



# Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans

Maanasa Raghavan<sup>1\*</sup>, Pontus Skoglund<sup>2\*</sup>, Kelly E. Graf<sup>3</sup>, Mait Metspalu<sup>4,5,6</sup>, Anders Albrechtsen<sup>7</sup>, Ida Moltke<sup>7,8</sup>, Simon Rasmussen<sup>9</sup>, Thomas W. Stafford Jr<sup>1,10</sup>, Ludovic Orlando<sup>1</sup>, Ene Metspalu<sup>6</sup>, Monika Karmin<sup>4,6</sup>, Kristiina Tambets<sup>4</sup>, Siiri Rootsi<sup>4</sup>, Reedik Mägi<sup>11</sup>, Paula F. Campos<sup>1</sup>, Elena Balanovska<sup>12</sup>, Oleg Balanovsky<sup>12,13</sup>, Elza Khusnutdinova<sup>14,15</sup>, Sergey Litvinov<sup>4,14</sup>, Ludmila P. Osipova<sup>16</sup>, Sardana A. Fedorova<sup>17</sup>, Mikhail I. Voevoda<sup>16,18</sup>, Michael DeGiorgio<sup>5</sup>, Thomas Sicheritz-Ponten<sup>9,19</sup>, Søren Brunak<sup>9,19</sup>, Svetlana Demeshchenko<sup>20</sup>, Toomas Kivisild<sup>4,21</sup>, Richard Villems<sup>4,6,22</sup>, Rasmus Nielsen<sup>5</sup>, Mattias Jakobsson<sup>2,23</sup> & Eske Willerslev<sup>1</sup>

**The origins of the First Americans remain contentious. Although Native Americans seem to be genetically most closely related to east Asians<sup>1–3</sup>, there is no consensus with regard to which specific Old World populations they are closest to<sup>4–8</sup>. Here we sequence the draft genome of an approximately 24,000-year-old individual (MA-1), from Mal'ta in south-central Siberia<sup>9</sup>, to an average depth of 1×. To our knowledge this is the oldest anatomically modern human genome reported to date. The MA-1 mitochondrial genome belongs to haplogroup U, which has also been found at high frequency among Upper Palaeolithic and Mesolithic European hunter-gatherers<sup>10–12</sup>, and the Y chromosome of MA-1 is basal to modern-day western Eurasians and near the root of most Native American lineages<sup>5</sup>. Similarly, we find autosomal evidence that MA-1 is basal to modern-day western Eurasians and genetically closely related to modern-day Native Americans, with no close affinity to east Asians. This suggests that populations related to contemporary western Eurasians had a more north-easterly distribution 24,000 years ago than commonly thought. Furthermore, we estimate that 14 to 38% of Native American ancestry may originate through gene flow from this ancient population. This is likely to have occurred after the divergence of Native American ancestors from east Asian ancestors, but before the diversification of Native American populations in the New World. Gene flow from the MA-1 lineage into Native American ancestors could explain why several crania from the First Americans have been reported as bearing morphological characteristics that do not resemble those of east Asians<sup>2,13</sup>. Sequencing of another south-central Siberian, Afontova Gora-2 dating to approximately 17,000 years ago<sup>14</sup>, revealed similar autosomal genetic signatures as MA-1, suggesting that the region was continuously occupied by humans throughout the Last Glacial Maximum. Our findings reveal that western Eurasian genetic signatures in modern-day Native Americans derive not only from post-Columbian admixture, as commonly thought, but also from a mixed ancestry of the First Americans.**

In 2009 we visited Hermitage State Museum in St. Petersburg, Russia, and sampled skeletal remains of a juvenile individual (MA-1) from the

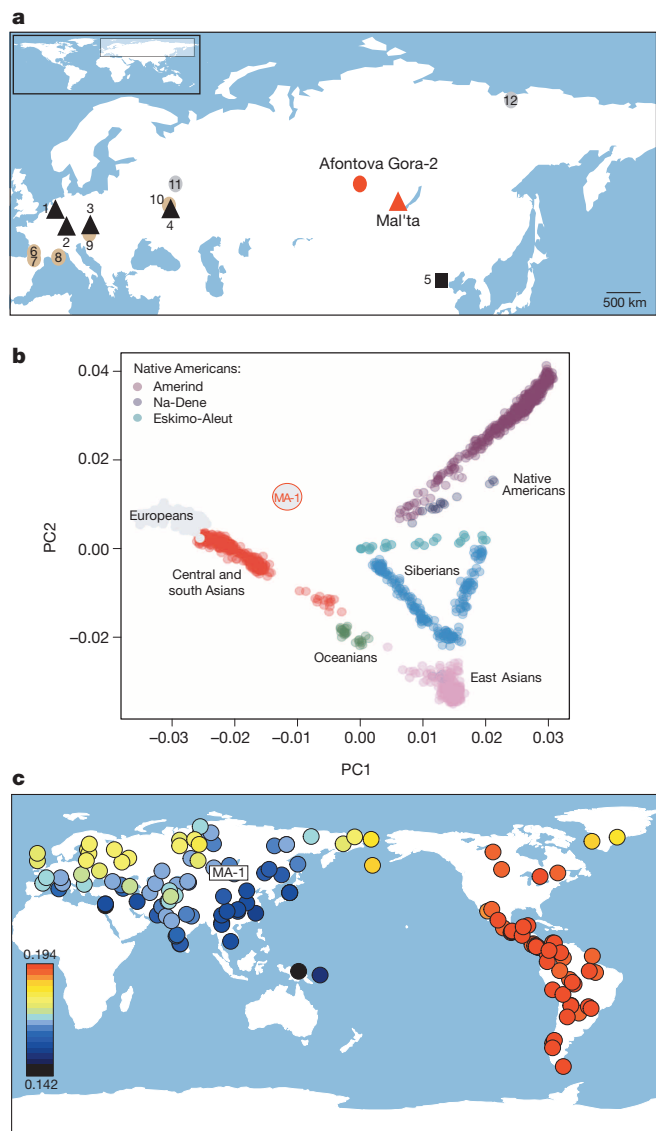
Mal'ta Upper Palaeolithic site in south-central Siberia. Mal'ta, located along the Belaya River near Lake Baikal, was excavated between 1928 and 1958 (ref. 9) and yielded a plethora of archaeological finds including 30 anthropomorphic Venus figurines, which are rare for Siberia but found at a number of Upper Palaeolithic sites across western Eurasia<sup>15–17</sup> (Fig. 1a and Supplementary Information, section 1). Accelerator mass spectrometry (AMS) <sup>14</sup>C dating of MA-1 produced an age of 20,240 ± 60 <sup>14</sup>C years before present or 24,423–23,891 calendar years before present (cal. BP) (Supplementary Information, section 2).

DNA from 0.15 g of bone from MA-1 was sequenced to an average depth of 1× (Supplementary Information, section 3). From one library (referred to as MA-1\_1<sup>st</sup> extraction in Supplementary Information, section 3.1), approximately 17% of the total reads generated mapped uniquely to the human genome, in agreement with good DNA preservation (see Supplementary Information Table 2). Low contamination rates were inferred for both mitochondrial DNA (mtDNA) (1.1%) and the X chromosome (1.6 to 2%; MA-1 is male) (Supplementary Information, section 5). The overall error rate for the data set was estimated to be 0.27%, with the most dominant errors being transitions typical of ancient DNA damage deriving from post-mortem deamination of cytosine<sup>18</sup> (Supplementary Information, section 6.1).

Phylogenetic analysis of the MA-1 mtDNA genome (76.6×) places it within mtDNA haplogroup U without affiliation to any known subclades, implying a lineage that is rare or extinct in sampled modern populations (Supplementary Information, section 7 and Supplementary Fig. 4a). Present-day distribution of haplogroup U encompasses a large area including North Africa, the Middle East, south and central Asia, western Siberia and Europe (Supplementary Fig. 4b), although it is rare or absent east of the Altai Mountains; that is, in populations living in the region surrounding Mal'ta. Haplogroup U has also been found at high frequency (>80%) in ancient hunter-gatherers from Upper Palaeolithic and Mesolithic Europe<sup>10–12</sup>. Our result therefore suggests a connection between pre-agricultural Europe and Upper Palaeolithic Siberia. The Y chromosome of MA-1 was sequenced to an average depth of 1.5×, with coverage across 5.8 million bases. Acknowledging

<sup>1</sup>Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5–7, 1350 Copenhagen, Denmark. <sup>2</sup>Department of Evolutionary Biology, Uppsala University, Norbyvägen 18D, Uppsala 752 36, Sweden. <sup>3</sup>Center for the Study of the First Americans, Texas A&M University, TAMU-4352, College Station, Texas 77845-4352, USA. <sup>4</sup>Estonian Biocentre, Evolutionary Biology group, Tartu 51010, Estonia. <sup>5</sup>Department of Integrative Biology, University of California, Berkeley, California 94720, USA. <sup>6</sup>Department of Evolutionary Biology, University of Tartu, Tartu 51010, Estonia. <sup>7</sup>The Bioinformatics Centre, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, Copenhagen 2200, Denmark. <sup>8</sup>Department of Human Genetics, The University of Chicago, Chicago, Illinois 60637, USA. <sup>9</sup>Center for Biological Sequence Analysis, Technical University of Denmark, Kongens Lyngby 2800, Denmark. <sup>10</sup>AMS <sup>14</sup>C Dating Centre, Department of Physics and Astronomy, University of Aarhus, Ny Munkegade 120, Aarhus DK-8000, Denmark. <sup>11</sup>Estonian Genome Center, University of Tartu, Tartu 51010, Estonia. <sup>12</sup>Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moskvorechye Street 1, Moscow 115479, Russia. <sup>13</sup>Vavilov Institute of General Genetics, Russian Academy of Sciences, Gubkina Street 3, Moscow 119991, Russia. <sup>14</sup>Institute of Biochemistry and Genetics, Ufa Scientific Centre, Russian Academy of Sciences, Ufa, Bashkortostan 450054, Russia. <sup>15</sup>Biology Department, Bashkir State University, Ufa, Bashkortostan 450074, Russia. <sup>16</sup>The Institute of Cytology and Genetics, Center for Brain Neurobiology and Neurogenetics, Siberian Branch of the Russian Academy of Sciences, Lavrentyeva Avenue, Novosibirsk 630090, Russia. <sup>17</sup>Department of Molecular Genetics, Yakut Research Center of Complex Medical Problems, Russian Academy of Medical Sciences and North-Eastern Federal University, Yakutsk, Sakha (Yakutia) 677010, Russia. <sup>18</sup>Institute of Internal Medicine, Siberian Branch of the Russian Academy of Medical Sciences, Borisa Bogatkova 175/1, Novosibirsk 630089, Russia. <sup>19</sup>Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby 2800, Denmark. <sup>20</sup>The State Hermitage Museum, 2, Dvortsovaya Ploshchad, St. Petersburg 190000, Russia. <sup>21</sup>Department of Biological Anthropology, University of Cambridge, Cambridge CB2 1QH, UK. <sup>22</sup>Estonian Academy of Sciences, Tallinn 10130, Estonia. <sup>23</sup>Science for Life Laboratory, Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden.

\*These authors contributed equally to this work.



**Figure 1 | Sample locations and MA-1 genetic affinities.** **a**, Geographical locations of Mal'ta and Afontova Gora-2 in south-central Siberia. For reference, Palaeolithic sites with individuals belonging to mtDNA haplogroup U are shown (red and black triangles): 1, Oberkassel; 2, Hohle Fels; 3, Dolni Vestonice; 4, Kostenki-14. A Palaeolithic site with an individual belonging to mtDNA haplogroup B is represented by the square: 5, Tianyuan Cave. Notable Palaeolithic sites with Venus figurines are marked by brown circles: 6, Laussel; 7, Lespugue; 8, Grimaldi; 9, Willendorf; 10, Gargarino. Other notable Palaeolithic sites are shown by grey circles: 11, Sungir; 12, Yana RHS. **b**, PCA (PC1 versus PC2) of MA-1 and worldwide human populations for which genomic tracts from recent European admixture in American and Siberian populations have been excluded<sup>19</sup>. **c**, Heat map of the statistic  $f_3(\text{Yoruba}; \text{MA-1}, X)$  where  $X$  is one of 147 worldwide non-African populations (standard errors shown in Supplementary Fig. 21). The graded heat key represents the magnitude of the computed  $f_3$  statistics.

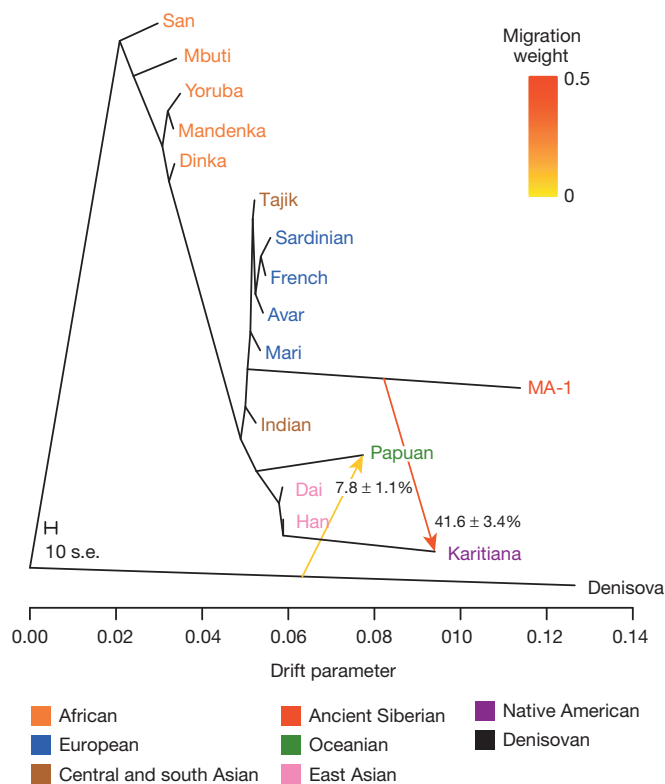
the low depth of coverage, we determined the most likely phylogenetic affiliation of the MA-1 Y chromosome to a basal lineage of haplogroup R (Supplementary Information, section 8 and Supplementary Fig. 5a). The extant sub-lineages of haplogroup R show regional spread patterns within western Eurasia, south Asia and also extend to the Altai region in southern Siberia (Supplementary Fig. 5b). The sister lineage to these extant sub-lineages of haplogroup R, haplogroup Q, is the most common haplogroup in Native Americans<sup>5</sup> and it was recently shown that, in Eurasia, haplogroup Q lineages closest to Native Americans are found in southern Altai<sup>7</sup>.

To get an overview of the genomic signature of MA-1, we conducted principal component analysis (PCA) using a large data set from worldwide human populations for which genomic tracts of recent European admixture in American and Siberian populations have been excluded<sup>19</sup> (Supplementary Information, section 10). In the first two principal components, MA-1 is intermediate between modern western Eurasians and Native Americans, but distant from east Asians (Fig. 1b). To investigate the relationship of MA-1 to global human populations in further detail, we used the  $f$ -statistics framework<sup>20</sup> to compute an 'outgroup'  $f_3$ -statistic, which is expected to be proportional to the amount of shared genetic history between MA-1 and each of 147 non-African populations from a large worldwide human single-nucleotide polymorphism (SNP) array data set (see Supplementary Information, section 14.2 for details on the  $f_3$ -statistics). We find that genetic affinity to MA-1 is greatest in two regions: first, the Americas; and second, northeast Europe and northwest Siberia, with north-to-south latitudinal clines in shared drift with MA-1 in both Europe and Asia (Fig. 1c and Supplementary Figs 21 and 22). Notably, the lack of genetic affinity between MA-1 and most populations in south-central Siberia today suggests that there was substantial gene flow into the region after the Last Glacial Maximum (LGM), mostly probably from east Asian sources (Supplementary Information, section 9.1.3).

We reconstructed admixture graphs using TreeMix<sup>21</sup> to relate the population history of MA-1 to 11 modern genomes from worldwide populations<sup>22</sup>, 4 new genomes from Eurasia (Mari, Avar, Indian and Tajik ancestry) and the Denisova genome<sup>22</sup> (Supplementary Information, section 11). The maximum-likelihood population tree inferred without admixture events places MA-1 on a branch that is basal to western Eurasians (Supplementary Fig. 12). However, a significant residual was observed between the empirical covariance for MA-1 and Karitiana, a Native American population, and the covariance predicted by the tree model (Supplementary Fig. 12). Consequently, gene flow between these lineages was inferred in all graphs incorporating two or more migration events (Fig. 2 and Supplementary Fig. 13). Bootstrap support for the migration edge from MA-1 to Karitiana, rather than from Karitiana to MA-1, was 99% in this analysis.

We investigated further the population history of MA-1 by conducting sequence read-based  $D$ -statistic tests<sup>23</sup> on proposed tree-like histories comprising MA-1 and combinations of 11 modern genomes (Supplementary Information, section 13). In agreement with the TreeMix results, these tests reject the tree  $((X, \text{Han}), \text{MA-1})$  where  $X$  represents Avar, French, Indian, Mari, Sardinian and Tajik, consistent with the MA-1 lineage sharing more recent ancestry with the western Eurasian branch after the split of Europeans and east Asians (Supplementary Table 13). This result also holds true when the Han Chinese is replaced with Dai, another east Asian population (Supplementary Table 13). Notably, we can also reject the tree  $((\text{Han}, \text{Karitiana}), \text{MA-1})$  ( $Z = 10.8$ ), suggesting gene flow between MA-1 and ancestral Native Americans, in accordance with the admixture graphs (Supplementary Table 13). This result is consistent with allele frequency-based  $D$ -statistic tests<sup>20</sup> on SNP arrays for 48 Native American populations of entirely First American ancestry<sup>19</sup>, indicating that all tested populations are equally related to MA-1 and that the admixture event occurred before the population diversification of the First American gene pool (Fig. 3a, Supplementary Information, section 14.4 and Supplementary Fig. 24).

The genetic affinity between Native Americans and MA-1 could be explained by gene flow after the split between east Asians and Native Americans, either from the MA-1 lineage into Native American ancestors or from Native American ancestors to the ancestors of MA-1. However, MA-1, at approximately 24,000 cal. BP, pre-dates time estimates of the Native American–east Asian population divergence event<sup>24,25</sup>. This presents little time for the formation of a diverged Native American gene pool that could have contributed ancestry to MA-1, suggesting gene flow from the MA-1 lineage into Native American ancestors. Such gene flow should also be detectable using modern-day western Eurasian populations in place of MA-1. Consistent with this,  $D$ -statistic tests



**Figure 2 | Admixture graph for MA-1 and 16 complete genomes.** An admixture graph with two migration edges (depicted by arrows) was fitted using TreeMix<sup>21</sup> to relate MA-1 to 11 modern genomes from worldwide populations<sup>22</sup>, 4 modern genomes produced in this study (Avar, Mari, Indian and Tajik), and the Denisova genome<sup>22</sup>. Trees without migration, graphs with different number of migration edges, and residual matrices are shown in Supplementary Information, section 11. The drift parameter is proportional to  $2N_e$  generations, where  $N_e$  is the effective population size. The migration weight represents the fraction of ancestry derived from the migration edge. The scale bar shows ten times the average standard error (s.e.) of the entries in the sample covariance matrix. Note that the length of the branch leading to MA-1 is affected by this ancient genome being represented by haploid genotypes.

estimated from outgroup-ascertained SNP data<sup>20</sup> reveal significant evidence ( $Z > 3$ ) for Middle Eastern, European, central Asian and south Asian populations being closer to Karitiana than to Han Chinese<sup>20</sup> (Fig. 3b and Supplementary Information, section 14.5). Similar signals were also observed when we replaced modern-day Han Chinese with data from chromosome 21 from a 40,000-year-old east Asian individual (Tianyuan Cave, China), which has been found to be ancestral to modern-day Asians and Native Americans<sup>26</sup> (Supplementary Information, section 14.5). Thus, if the gene flow direction was from Native Americans into western Eurasians it would have had to spread subsequently to European, Middle Eastern, south Asian and central Asian populations, including MA-1 before 24,000 years ago. Moreover, as Native Americans are closer to Han Chinese than to Papuans (Fig. 3c), Native American-related gene flow into the ancestors of MA-1 is expected to result in MA-1 also being closer to Han Chinese than to Papuans. However, our results suggest that this is not the case ( $D(\text{Papuan}, \text{Han}; \text{Sardinian}, \text{MA-1}) = -0.002 \pm 0.005$  ( $Z = -0.36$ )), which is compatible with all or almost all of the gene flow being into Native Americans (Supplementary Information, section 14.6). Similar results are obtained when MA-1 is replaced with most modern-day western Eurasian populations, except populations with recent admixture from east Asia (Russian, Adygei and Burusho) and Africa (Middle Eastern populations) (Fig. 3c). The most parsimonious explanation for these results is that Native Americans have mixed origins, resulting from admixture between peoples related to modern-day east Asians and western Eurasians. Admixture graphs fitted with MixMapper<sup>27</sup> model Karitiana as having 14–38%

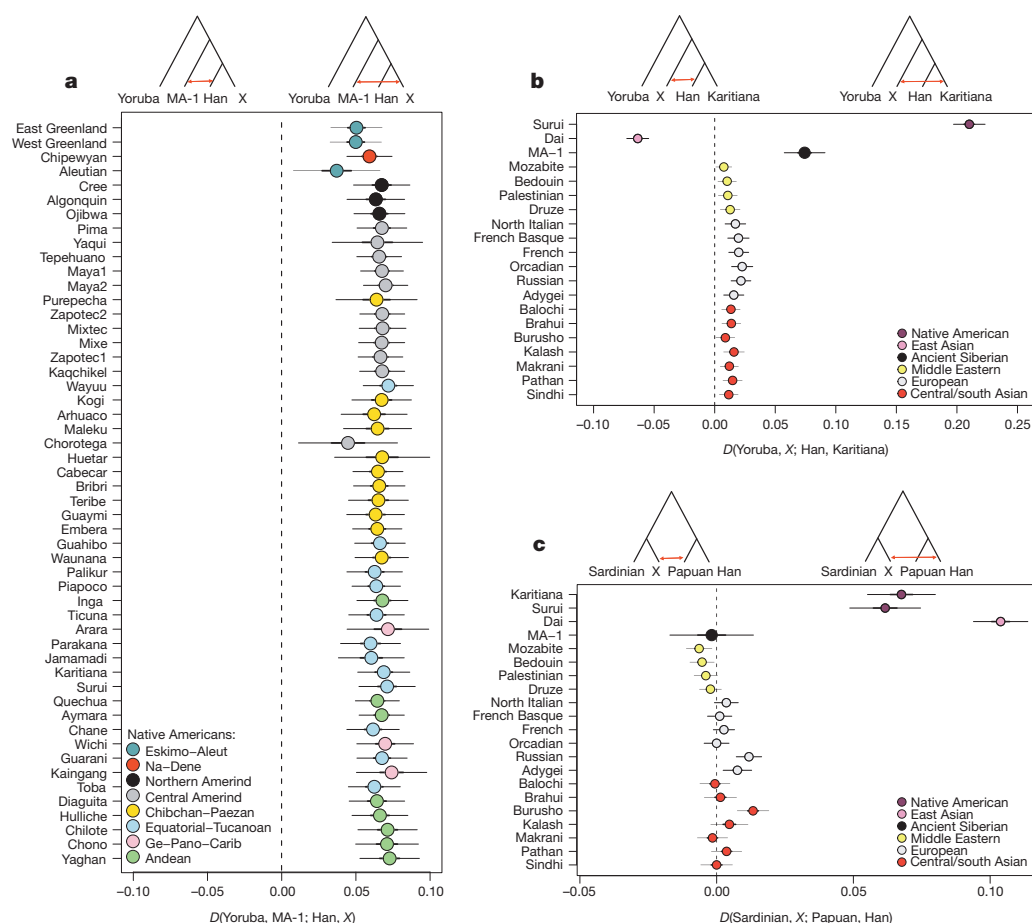
western Eurasian ancestry and 62–86% east Asian ancestry, but we caution that these estimates assume unadmixed ancestral populations (Supplementary Information, section 12).

Importantly, in addition to the low contamination rates and rare or extinct uniparental lineages, we exclude modern DNA contamination as being the source of the observed population affinities of MA-1 for three reasons. First, we corrected the sequence read-based  $D$ -statistics tests for differing amounts of contamination, using a European individual as the contamination source (Supplementary Information, section 13.5). We find similar outcomes for corrected and uncorrected tests (Supplementary Fig. 20), even when contamination levels larger than that estimated for MA-1 are considered, confirming that our results are not affected by contamination from a European source. Second, restricting the PCA to sequences with evidence of post-mortem degradation gives results that are comparable with those using the complete data set (Supplementary Information, section 15). Finally, the genome sequence of the researcher (Indian ancestry) who carried out DNA extraction and library preparation of MA-1 enables us to exclude the researcher as a source of contamination (Supplementary Information, sections 11 and 13). In addition, we exclude post-Columbian European admixture (after 1492 AD) as an explanation for the genetic affinity between MA-1 and Native Americans for three reasons. First, for SNP array-based analyses, we take recent European admixture into account by using a data set masked for inferred admixed genomic regions<sup>19</sup>. Second, allele frequency-based  $D$ -statistic tests<sup>20</sup> show that all 48 tested modern-day populations with First American ancestry<sup>19</sup> are equally related to MA-1 within the resolution of our data (Supplementary Information, section 14.4), which would not be expected if the signal was driven by recent European admixture. Third, MA-1 is closer to Native Americans than any of the 15 tested European populations (Supplementary Information, section 14.8).

Human dispersals in northeast Asia immediately before and after the LGM are most likely to have led to the settlement of Beringia, and ultimately the Americas<sup>28</sup>. As MA-1 pre-dates the LGM, we investigated whether the genetic composition of southern Siberia changed during the LGM by generating a low-coverage data set ( $\sim 0.1\times$ ) of a post-LGM individual from Afontova Gora-2 (AG-2) (ref. 14), located on the western bank of the Enisei River in south-central Siberia (Fig. 1a). We obtained a direct AMS  $^{14}\text{C}$  date of  $13,810 \pm 35$   $^{14}\text{C}$  years before present or 17,075–16,750 cal. BP for AG-2 (Supplementary Information, section 2). Despite substantial present-day DNA contamination in this sample (Supplementary Information, section 5), we find that AG-2 shows close similarity to the genetic profile of MA-1 on a PCA (Supplementary Information, section 15 and Supplementary Fig. 29) and is significantly closer to Karitiana than to Han ( $D(\text{Yoruba}, \text{AG-2}; \text{Han}, \text{Karitiana}) = 0.078 \pm 0.004$ ,  $Z = 19.9$ ) (Supplementary Information, section 15). We observe consistent results when restricting analyses to sequences with evidence of post-mortem degradation (Supplementary Information, section 15 and Supplementary Fig. 29), implying that southern Siberia may have experienced genetic continuity through the environmentally harsh LGM.

Our study has four important implications. First, we find evidence that contemporary Native Americans and western Eurasians share ancestry through gene flow from a Siberian Upper Palaeolithic population into First Americans. Second, our findings may provide an explanation for the presence of mtDNA haplogroup X in Native Americans, which is related to western Eurasians but not found in east Asian populations<sup>29</sup>. Third, such an easterly presence in Asia of a population related to contemporary western Eurasians provides a possibility that non-east Asian cranial characteristics of the First Americans<sup>13</sup> derived from the Old World via migration through Beringia, rather than by a trans-Atlantic voyage from Iberia as proposed by the Solutrean hypothesis<sup>30</sup>. Fourth, the presence of an ancient western Eurasian genomic signature in the Baikal area before and after the LGM suggests that parts of south-central Siberia were occupied by humans throughout the coldest stages of the last ice age.





**Figure 3 | Evidence of gene flow from a population related to MA-1 and western Eurasians into Native American ancestors.** Allele frequency-based  $D$ -statistic tests<sup>20</sup> of the forms **a**,  $D(\text{Yoruba}, \text{MA-1}; \text{Han}, \text{X})$ , where  $X$  represents modern-day populations from North and South America. The  $D$ -statistic is significantly positive for all the tests, providing evidence for gene flow between Native American ancestors and the MA-1 population lineage; however, it is not informative with respect to the direction of gene flow. **b**,  $D(\text{Yoruba}, \text{X}; \text{Han}, \text{Karitiana})$ , where  $X$  represents non-African populations. Since all of the 17 tested western Eurasian populations are closer to Karitiana than to Han Chinese, the most parsimonious explanation is that Native Americans have western Eurasian-related ancestry. **c**,  $D(\text{Sardinian}, \text{X}; \text{Papuan}, \text{Han})$ , where  $X$  represents non-African populations. MA-1 is not significantly closer to Han Chinese than to Papuans, which is compatible with MA-1 having no Native American-related admixture in its ancestry. Thick and thin error bars correspond to 1 and 3 standard errors of the  $D$ -statistic, respectively.

## METHODS SUMMARY

A humerus from Mal'ta (MA-1) and a humerus from Afontova Gora-2 (AG-2) were sampled at the Hermitage Museum, St. Petersburg, for ancient DNA and radiocarbon dating analyses. Informed consent and institutional approval were obtained for the genome sequencing of modern individuals (Avar, Mari, Tajik and Indian ancestry) in accordance with human demographic studies, with ethical approval from The National Committee on Health Research Ethics, Denmark (H-3-2012-FSP21). Ancient and modern DNA extracts were built into Illumina libraries and sequenced on the Illumina HiSeq and MiSeq platforms (Supplementary Information, section 3). Reads were mapped to the human reference genome builds hg18 and 37.1 (Supplementary Information, section 4). Principal component analysis was carried out to investigate affinities to modern populations (Supplementary Information, sections 10 and 15). Admixture graphs were fitted to the observed allele frequencies using TreeMix<sup>21</sup> (Supplementary Information, section 11) and MixMapper<sup>27</sup> (Supplementary Information, section 12). Tree-like population models were tested using  $D$ -statistics based on both sequence read data<sup>23</sup> (Supplementary Information, section 13) and allele frequency data from SNP arrays<sup>20</sup> (Supplementary Information, section 14.3). Shared genetic drift with modern populations was estimated using  $f_3$ -statistics<sup>20</sup> (Supplementary Information, section 14).

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 July; accepted 4 October 2013.

Published online 20 November 2013.

- Turner, C. G. Advances in the dental search for native american origins. *Acta Anthropogenet.* **8**, 23–78 (1984).
- Hubbe, M., Harvati, K. & Neves, W. Paleoamerican morphology in the context of European and East Asian Pleistocene variation: implications for human dispersion into the New World. *Am. J. Phys. Anthropol.* **144**, 442–453 (2011).
- Schurr, T. The peopling of the New World: perspectives from molecular anthropology. *Annu. Rev. Anthropol.* **33**, 551–583 (2004).
- O'Rourke, D. H. & Raff, J. A. The human genetic history of the Americas: the final frontier. *Curr. Biol.* **20**, R202–R207 (2010).
- Lell, J. T. *et al.* The dual origin and siberian affinities of native american Y chromosomes. *Am. J. Hum. Genet.* **70**, 192–206 (2002).

- Starikovskaya, E. B. *et al.* Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. *Ann. Hum. Genet.* **69**, 67–89 (2005).
- Dulik, M. C. *et al.* Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native American and Indigenous Altaians. *Am. J. Hum. Genet.* **90**, 229–246 (2012).
- Regueiro, M., Alvarez, J., Rowold, D. & Herrera, R. J. On the origins, rapid expansion and genetic diversity of Native Americans from hunting-gatherers to agriculturalists. *Am. J. Phys. Anthropol.* **150**, 333–348 (2013).
- Gerasimov, M. M. in *The Archaeology and Geomorphology of Northern Asia: Selected Works 5–32* (University of Toronto Press, 1964).
- Bramanti, B. *et al.* Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326**, 137–140 (2009).
- Malmström, H. *et al.* Ancient DNA reveals lack of continuity between Neolithic hunter-gatherers and contemporary Scandinavians. *Curr. Biol.* **19**, 1758–1762 (2009).
- Fu, Q. *et al.* A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* **23**, 553–559 (2013).
- Owsley, D. W. & Jantz, R. L. in *Claiming the Stones-Naming the Bones: Cultural Property and the Negotiation of National and Ethnic Identity* (Getty Research Institute, 2002).
- Astakhov, S. N. *Paleolit Eniseia: Paleoliticheskie Stoianki Afontovo Gore v G. Krasnoirske* (Evropaikii Dom, 1999).
- Gamble, C. Interaction and alliance in Palaeolithic society. *Man (Lond)* **17**, 92–107 (1982).
- Abramova, Z. *L'art Paléolithique d'Europe Orientale et de Sibérie* (Jérôme Millon, 1995).
- White, R. The women of Brassempouy: a century of research and interpretation. *J. Archaeol. Method and Theory* **13**, 250–303 (2006).
- Hansen, A. J., Willerslev, E., Wiuf, C., Mourier, T. & Arctander, P. Statistical evidence for miscoding lesions in ancient DNA templates. *Mol. Biol. Evol.* **18**, 262–265 (2001).
- Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
- Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
- Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
- Meyer, M. *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226 (2012).
- Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).



24. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
25. Wall, J. D. *et al.* Genetic variation in Native Americans, inferred from latino SNP and resequencing data. *Mol. Biol. Evol.* **28**, 2231–2237 (2011).
26. Fu, Q. *et al.* DNA analysis of an early modern human from Tianyuan Cave, China. *Proc. Natl Acad. Sci. USA* **110**, 2223–2227 (2013).
27. Lipson, M. *et al.* Efficient moment-based inference of admixture parameters and sources of gene flow. *Mol. Biol. Evol.* (2013).
28. Goebel, T. Pleistocene human colonization of Siberia and peopling of the Americas: an ecological approach. *Evol. Anthropol.* **8**, 208–227 (1999).
29. Brown, M. D. *et al.* mtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? *Am. J. Hum. Genet.* **63**, 1852–1861 (1998).
30. Bradley, B. & Stanford, D. The North Atlantic ice-edge corridor: a possible Palaeolithic route to the New World. *World Archaeol.* **36**, 459–478 (2004).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank the Hermitage State Museum for providing access to the Mal'ta and Afontova Gora-2 human remains. We also thank the Danish National High-Throughput DNA Sequencing Centre and T. Reisberg for technical assistance. This work was supported by the Danish National Research Foundation and the Lundbeck Foundation (E.W. and M.R.) and the Arctic Social Sciences Program, National Science Foundation (grant PLR-1003725 to K.E.G.). R.V., M.M., M.K., E.M., K.T., S.Ro. and R.M. were supported by the European Regional Development Fund (European Union) through the Centre of Excellence in Genomics to Estonian Biocentre and University of Tartu and Estonian Basic Research grant SF0270177As08. M.M. thanks the Estonian Science Foundation grant no. 8973 and Baltic-American Freedom Foundation Research Scholarship program and M.I.V. thanks the Government of Russian Federation grant no. 14.B25.31.0033 (to E. I. Rogaev). M.D. was supported by the US National Science Foundation (grant DBI-1103639). Computational analyses were carried out at the High Performance Computing Center, University of Tartu, and the Swedish National Infrastructure for Computing (SNIC-UPPMAX, project b2012063).

**Author Contributions** E.W. and K.E.G. conceived the project. E.W. headed the project. E.W. and M.R. designed the experimental research project setup. S.D. and K.E.G. provided access to the Mal'ta and Afontova Gora-2 samples, and K.E.G. provided

archaeological context for the samples. T.W.S. Jr performed AMS dating. E.B. and O.B. (Tajik individual), E.K. and S.L. (Mari and Avar individuals) provided modern DNA extracts for complete genome sequencing. E.K. and S.L. (Kazakh, Kirghiz, Uzbek and Mari individuals), L.P.O. (Selkup individuals), S.A.F. (Even, Dolgan and Yakut individuals) and M.I.V. (Altai individuals) provided access to modern DNA extracts for genotyping. R.V. carried out Illumina chip analysis on modern samples. P.F.C. performed DNA extraction from the Indian individual. M.R. performed the ancient extractions and library constructions on the modern and ancient samples — the latter with input from L.O. M.R. coordinated the sequencing. M.R. and S.Ra. performed mapping of MA-1 and AG-2 data sets with input from L.O. S.Ra., T.S.-P. and S.B. provided super-computing resources, developed the next-generation sequencing pipeline and performed mapping and genotyping for all the modern genomes. M.R. performed DNA damage analysis with input from L.O. M.M. performed the admixture analysis. M.M., E.M., K.T. and R.V. performed the mtDNA analysis. M.M., M.K., S.Ro., T.K., R.V. and R.M. performed the Y-chromosome analysis. A.A. and I.M. performed the autosomal contamination estimates, error rate estimates, *D*-statistics tests based on sequence reads and ngsAdmix analyses. P.S. performed biological sexing, mtDNA contamination estimates, PCA, TreeMix, MixMapper, *D*-statistic tests based on allele frequencies, *f*<sub>3</sub>-statistics and phenotypic analyses, and analysis of AG-2 using nucleotide misincorporation patterns under the supervision of R.N. and M.J. M.R., P.S. and E.W. wrote the majority of the manuscript with critical input from R.N., M.J., M.M., K.E.G., A.A., I.M. and M.D. M.M., A.A. and I.M. contributed equally to this work.

**Author Information** Sequence data for MA-1 and AG-2, produced in this study, are available for download through NCBI SRA accession number SRP029640. Data from the Illumina genotyping analysis generated in this study are available through GEO Series accession number GSE50727; PLINK files can be accessed from [http://www.ebc.ee/free\\_data](http://www.ebc.ee/free_data). In addition, the above data and alignments for the published modern genomes, Denisova genome, Tianyuan individual and the two ancient genomes are available at <http://www.cbs.dtu.dk/suppl/malta>. Raw reads and alignments for the four modern genomes sequenced in this study are available for demographic research under data access agreement with E.W. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.W. ([ewillerslev@snm.ku.dk](mailto:ewillerslev@snm.ku.dk)).

## METHODS

**Samples.** A humerus (MA-1) from Mal'ta and a humerus (AG-2) from Afontova Gora-2 were sampled at the Hermitage Museum, St. Petersburg, Russia in 2009 for ancient DNA analysis and accelerator mass spectrometry (AMS)  $^{14}\text{C}$  dating. In addition, four modern human samples (Avar, Mari, Tajik and Indian) were obtained for genome sequencing in accordance with informed consent requirements for human demographic studies. Ethical approval for genome sequencing of the above four modern samples was acquired from The National Committee on Health Research Ethics, Denmark (H-3-2012-FSP21).

**Radiocarbon dating.** AMS  $^{14}\text{C}$  dating was carried out on the two ancient bone samples following standard protocols<sup>31,32</sup> (Supplementary Information, section 2). Contemporary  $^{14}\text{C}$  standards included National Bureau of Standards Oxalic Acid-I and ANU sucrose. Respective chemistry and combustion backgrounds were determined by using >70,000-year-old collagen isolated from the fossil *Eschrichtius robustus* (grey whale)<sup>32,33</sup> and Sigma Aldrich L-Alanine (catalogue number A7627). The graphitized samples and standards were analysed at the University of California-Irvine WM Keck Carbon Cycle Accelerator Mass Spectrometry Laboratory (UCIAMS). The  $^{14}\text{C}$  dates were calibrated using OxCal 4.2 (ref. 34) and the INTCAL09 data set<sup>35</sup>.

**Genome sequencing and read processing.** DNA extractions and library constructions for the ancient samples were performed in a laboratory facility dedicated to the analysis of ancient DNA (Centre for GeoGenetics, Copenhagen). Bone powder from MA-1 and AG-2 (149 mg and 119 mg, respectively) was extracted using a silica spin-column protocol<sup>11,36,37</sup> (Supplementary Information, section 3.1.1). Undigested pellets were subject to another round of digestion. Blood samples from one individual each of Avar, Mari and Tajik ancestry were extracted using standard protocol<sup>38</sup> (Supplementary Information, section 3.2.2). A saliva sample from an individual of Indian ancestry was extracted using a prepIT•L2P extraction kit (DNA Genotek) (Supplementary Information, section 3.2.2). Illumina libraries were constructed on the ancient and modern extracts (Supplementary Information, sections 3.1.2 and 3.2.3). The protocols outlined in the kit manuals (GS FLX Titanium Rapid Library Preparation Kit, 454 Life Sciences, Roche, Branford, CT and NEBNext DNA Sample Prep Master Mix Set 2, New England Biolabs, E6070) as well as in a previous paper<sup>39</sup> were followed. Equimolar pools of the ancient (100 cycles, single-read mode) and modern libraries (100 cycles, paired-end mode) were sequenced on the Illumina HiSeq 2000 at the Danish National High-Throughput DNA Sequencing Centre. The ancient libraries were sequenced to near-saturation.

Read processing was performed on the ancient and modern genomes produced in this study as well as previously published genomes (Supplementary Information, sections 4.1 and 4.2). The latter genomes included 11 high-coverage modern genomes<sup>22</sup>, one low-coverage Cambodian genome<sup>40</sup>, and the Denisovan<sup>22</sup> and Tianyuan<sup>26</sup> ancient data sets. All sequences were trimmed using AdapterRemoval<sup>41</sup> and mapped to the human reference genome builds hg18 and 37.1 using the Burrows-Wheeler Aligner (BWA)<sup>42</sup>. The seed length option was disabled for ancient reads to optimize the mapping efficiency<sup>43</sup>. Polymerase chain reaction (PCR) duplicates were removed using Picard MarkDuplicates (<http://picard.sourceforge.net>). All modern samples (except the Cambodian genome) and the Denisova individual were genotyped using samtools mpileup and bcftools<sup>44</sup>, and filtered to achieve a high-confidence SNP set (Supplementary Information, section 4.2). Only bi-allelic sites were included when producing the final call set and the individual calls were merged to a final set using Genome Analysis Toolkit (GATK) CombineVariants-2.5-2 (ref. 45).

**Contamination and error rate estimation.** Mitochondrial DNA (mtDNA) contamination rates for MA-1 and AG-2 were estimated by identifying consensus calls in the ancient mtDNA data set that are private or near-private to the ancient individual (at an allele frequency of less than 1% in a set of 311 modern human mtDNA genomes)<sup>46</sup> (Supplementary Information, section 5.1). The near-private consensus alleles and potential contaminating reads at these positions were counted, and a 95% confidence interval was obtained assuming that the allele observed in each read is a random outcome of drawing one of two alleles (endogenous and contaminant). Positions with a depth of less than 10 $\times$  were excluded, as were positions where the consensus allele was either C or G in a transition polymorphism, as these are sensitive to post-mortem nucleotide misincorporations. A phred-scaled base quality of 30 was required.

As we found both individuals (MA-1 and AG-2) to be males by comparing the number of alignments to the X and Y chromosomes<sup>47</sup> (Supplementary Information, section 4.3), it was possible to obtain X chromosome-based contamination estimates using previously published methods<sup>48</sup> (Supplementary Information, section 5.2). These estimates were based on a fixed set of SNPs known to be polymorphic in European HapMap phase II release 27 data<sup>49</sup>. This SNP data set was pruned such that polymorphic sites were more than 10 bases apart. The same HapMap data was used for estimating allele frequencies in Europeans. The MA-1

and AG-2 data sets were filtered to remove: regions homologous between the X and Y chromosomes; reads mapping non-uniquely to multiple regions of the genome with more than 98% identity; reads with mapping quality score less than 30 and base quality score less than 20; and sites with a read depth of less than 3 (or 2 depending on library depth) or above 40.

The error rates for the sequenced ancient and modern libraries were estimated using a method similar to a previously published method<sup>40</sup> that makes use of a high quality genome (Supplementary Information, section 6.1). The estimates were based on the rationale that any given human sample should have the same expected number of derived alleles compared to some outgroup, in this case the chimpanzee, panTro2, from the multiway alignment hg19 multiz46. The numbers of derived alleles were counted from the high-quality genome (individual NA06985 from the 1000 Genomes Project Consortium<sup>50</sup>) and the error rate estimates were based on the assumption that any excess of derived alleles (compared to the high quality genome) observed in our sample is due to errors. The overall error rates were estimated using a method of moment estimator, while the type specific error rates were estimated using a maximum likelihood approach. The model and the estimation methods are described in detail elsewhere<sup>39</sup>. All reads with a mapping quality score less than 30 and all bases with a base quality score less than 20 were excluded.

**mtDNA and Y-chromosome haplogroup determination.** Sequence reads from MA-1 were mapped to the revised Cambridge Reference Sequence (rCRS, NC\_012920.1) and filtered for PCR duplicates and paralogs, requiring a minimum mapping quality of 25 (Supplementary Information, section 4.1). A file of variants filtered for a minimum depth of 10, was generated (Supplementary Information, section 7). Indels were excluded from the analysis. mtDNA sequences from the individual Dolni Vestonice 14 (DV-14; GenBank accession number KC521458), basal to the extant mtDNA haplogroup U5 (ref. 12), was included in the analysis for comparison. Both the MA-1 and DV-14 mtDNA sequences were analysed for the presence of diagnostic mutations of the major sub-haplogroups of extant mtDNA haplogroup U lineages, using information from mtDNA tree Build 15 (Sept 30, 2012)<sup>51</sup>. A phylogenetic tree including all major extant branches of mtDNA haplogroup U was built, with the age estimates (kiloyears  $\pm$  s.d.) of the different sub-haplogroups<sup>52</sup> (Supplementary Fig. 4a). To show the present spread of haplogroup U and its different sub-haplogroups, the average frequencies, divided into four frequency classes, were calculated in regional groups, using a data set consisting of approximately 30,000 partial mtDNA genomes (references in Supplementary Information, section 7).

Owing to low depth of coverage of the MA-1 individual, genotyping at each site on the Y chromosome was performed by selecting the allele with the highest frequency of bases with a base quality of 13 or higher (Supplementary Information, section 8). A multi-fasta file was generated from the variable positions on the Y chromosomes available from 24 Complete Genomics public genomes<sup>53</sup>. SNPs were filtered for quality (using the threshold VQHIGH as defined by Complete Genomics), with tri-allelic positions excluded and only Y-chromosome regions determined as phylogenetically informative being used<sup>54</sup>. This yielded a final data set of 22,492 positions that was merged with MA-1 Y chromosome data. A neighbour joining tree with default parameters in MEGA phylogenetic software<sup>55</sup> was constructed (Supplementary Fig. 5a). Phylogenetically informative positions and their state in MA-1 were then determined to confirm the placement of MA-1 on the tree. Non-informative positions, including those with more than four Ns in the public data set, were excluded (633 positions). Moreover, the following positions were also excluded which were: in reference state in all individuals, including MA-1 (7,172 positions); N in MA-1 and either N or reference state among all individuals (9,682 positions); 'N-ref', those with only N or reference state among all individuals (586 positions), and 'N-alt', positions with alternative alleles, but difficult to classify (11 positions); 'reference-specific' (79 positions); and 'recurrent' (28 positions). This resulted in 4,301 positions being retained that were classified according to their haplogroup affiliations. Among those phylogenetically informative positions, 1,889 non-N positions were retrieved from MA-1.

**Principal component analysis.** A single read was sampled from each position in the MA-1 data set, which overlapped with SNPs in a data set compiled from a previous paper<sup>19</sup> in which the authors had used local ancestry inference to mask segments of European and African ancestry in Siberian and Native American populations<sup>56-59</sup> (Supplementary Information, section 10). A phred-scaled mapping quality of 30 and base quality score of 30 was required in the sequence data for a haploid genotype to be called, and reads with indels were excluded. SNPs with minor allele frequency of <1% in the total data set were removed. To reduce the effect of nucleotide misincorporations, the first and last three bases of each sequence read in the MA-1 data were excluded. SNPs where there was no information from MA-1 were excluded, and a single haploid genotype was randomly sampled from each modern individual to match the single-pass nature of the shotgun data<sup>60</sup>. PCA was performed on various population subsets separately using EIGENSOFT 4.0 (ref. 61), removing one SNP from each pair for which

linkage disequilibrium exceeded a low arbitrary threshold ( $r^2 > 0.2$ ). Transition SNPs, where the ancient individual displayed a T or an A<sup>62</sup>, as well as triallelic SNPs, were excluded.

To look more closely at the genetic affinities of AG-2 to modern-day populations, data from non-African populations<sup>59,63,64</sup> were used as a reference panel and PCA was performed as detailed above (Supplementary Information, section 15). To compare the PCA results from MA-1 and AG-1, Procrustes transformation was performed as described in a previous paper<sup>62</sup>, rotating the PC1–PC2 configurations obtained for the two individuals to the configuration obtained using only the reference panel (Supplementary Information, section 15). The analysis was repeated using only those sequences which displayed a C → T mismatch consistent with post-mortem ancient DNA nucleotide misincorporations (PMD) in the first five bases of the sequence read (requiring a base quality of at least 30) (Supplementary Information, section 15).

**Admixture graph inference.** To infer admixture graphs, a total of 17 individuals were used: the archaic Denisova genome<sup>22</sup>; 11 present-day individuals<sup>22</sup>; the 4 novel genomes from this study (Supplementary Information, section 4.2); and the MA-1 genome (Supplementary Information, section 11). Haploid genotypes from MA-1 were added to variants identified in the other individuals, as in the PCA analysis to alleviate the increased rate of errors in low-coverage ancient DNA sequence data. If multiple sequence reads overlapped a position, one read was randomly sampled<sup>23</sup>. This avoids biasing for, or against, heterozygotes and renders the MA-1 data haploid. All transition SNPs were excluded and MA-1 sequence reads with a mapping quality less than 30 and bases with base quality less than 30 were discarded. Positions at which there was no data from one of the individuals in the analysis were also excluded. This resulted in a final count of 156,250 SNPs for the main analysis. TreeMix<sup>21</sup> (version 1.12) was used to build ancestry graphs assuming 0 to 10 migration edges, the placement and weight of each being optimized by the algorithm. TreeMix was run using the ‘global’ option, which corresponds to performing a round of global rearrangements of the graph after initial fitting. Sample size correction was also disabled, as all the populations consisted of single individuals (‘-noss’). Standard errors were estimated in blocks with 500 SNPs in each. For those analyses that included one or more *a priori* specified events, a round of optimization was performed on the original migration edge (option ‘-climb’).

Admixture graphs relating MA-1 to modern groups were also inferred using MixMapper v1.0 (ref. 27) (Supplementary Information, section 12). A scaffold tree was constructed using four African genomes (San, Yoruba, Mandenka, Dinka), and Sardinian and Han<sup>22</sup> genomes, to which MA-1 and other genomes were fitted. All transitions were excluded, and standard errors of the *f*-statistics were estimated using 500 bootstrap replicates over 50 blocks of the autosomal genome.

**D-statistics.** To investigate the relationship between MA-1 and a number of modern populations, a sequence read-based *D*-statistic test (‘ABBA-BABA test’), equivalent to previously published tests<sup>23,40</sup>, was applied to sequencing data from a single genome from each of the populations of interest (Supplementary Information, section 13). MA-1 and 11 high-coverage present-day genomes were included in this test. For the chimpanzee outgroup, the multiway alignment, which includes both chimpanzee and human (panTro2 from the hg19 multiz46), was used. The data were filtered as follows before calculating the sequence read-based *D*-statistic<sup>39</sup>. First, all reads with mapping quality below 30 were removed. Subsequently, bases of low quality were removed by dividing all bases into eight base categories: A, C, G, T on the plus strand and A, C, G, T on the minus strand. The lowest-scoring 50% of bases from each of the eight categories were then discarded. More specifically, within each base category, we found the highest base quality score, *Q*, for which less than half of the bases in the base category had a quality score smaller than *Q*. We then removed all bases with quality score smaller than *Q*, and randomly sampled and removed bases with quality score equal to *Q* until 50% of the bases from the base category had been removed in total. The data were filtered separately for each of the eight base categories to avoid bias in the test in case of significant difference in the base quality between the categories. After filtering, a single base was sampled at each site for each individual in order to avoid introducing bias due to differences in sequencing depth. Finally, all sites containing transitions were removed. Based on the filtered data, *D*-statistics were calculated and to assess if these were significantly different from 0, standard errors and *Z* scores were obtained using a method known as ‘delete-m Jackknife for unequal *m*’, with a block size of 5 megabases<sup>65</sup>.

For genotype data from SNP arrays we computed an allele frequency-based *D*-statistic test, which is a generalization of the sequence read-based test (Supplementary Information, section 14.3). We used previously presented estimators<sup>20,66</sup>, obtaining standard errors using a block jackknife procedure over 5-megabase blocks in the genome, except for the tests with the Tianyuan data (chromosome 21), in which case we used 100-kb blocks to increase power. Two main data sets were used: first, a published SNP data set (364,470 SNPs) masked for European and

African ancestries in Siberian and Native American populations<sup>19</sup>, which was merged with additional data from Finnish populations<sup>63</sup>; and second, SNPs ascertained in San and Yoruban individuals and typed in worldwide populations<sup>20</sup>. As the San and Yoruba populations are approximate outgroups to non-African populations, this data are unbiased for all comparisons between non-Africans. Transition SNPs were included but the first and last three bases of each sequence read were excluded since the majority of nucleotide misincorporations occur at the ends of ancient DNA templates (Supplementary Information, section 6.2). For other tests, (in Supplementary Information, section 14), SNP data described in Supplementary Table 11 were used. We sampled a single read at each position from the MA-1 data as in the principal component analysis.

**Outgroup *f*<sub>3</sub>-statistics.** Classical measures of pairwise genetic distance, such as Wright’s fixation index *F*<sub>ST</sub>, are sensitive to genetic drift that has occurred since the divergence of the two test populations. If such lineage-specific genetic drift differs between populations that share an equal amount of genetic history with an ancient individual, the ancient individual would be observed as being closer to the modern populations with the least degree of historical genetic drift using distance-based methods such as *F*<sub>ST</sub>. To circumvent these issues and obtain a statistic that is informative of the genetic relatedness between a particular sample and each candidate population in a reference set, an ‘outgroup *f*<sub>3</sub>-statistic’ was computed (Supplementary Information, section 14.2). The expected value of the *f*<sub>3</sub>-statistic<sup>20</sup>, *f*<sub>3</sub>(Outgroup; A, B), equals the sum of expected squared change in allele frequency (normalized for heterozygosity in the outgroup) due to genetic drift on the path in the population tree from the outgroup to the root and from the root to the ancestor of populations A and B. As genetic drift in the lineage specific to the outgroup is expected to be constant regardless of which populations A and B are used (in the absence of gene flow), the remaining variation between statistics will depend on how much genetic history is shared between populations A and B. We used Yoruba as an outgroup to non-African populations and computed the statistic *f*<sub>3</sub>(Yoruba; MA-1, X) to investigate the shared history of MA-1 and a set of 147 worldwide candidate populations (as X) obtained by merging several data sets (Supplementary Figs 21 and 22), and we corroborated major patterns using SNPs from a San individual from southern Africa (Supplementary Information, section 14).

- Stafford, T. W. Jr, Jull, A. J. T., Brendel, K., Duhamel, R. & Donahue, D. Study of bone radiocarbon dating accuracy at the University of Arizona NSF accelerator facility for radioisotope analysis. *Radiocarbon* **29**, 24–44 (1987).
- Stafford, T. W. Jr, Brendel, K. & Duhamel, R. Radiocarbon, <sup>13</sup>C and <sup>15</sup>N analysis of fossil bone: removal of humates with XAD-2 resin. *Geochim. Cosmochim. Acta* **52**, 2257–2267 (1988).
- Stafford, T. W. Jr, Hare, P. E., Currie, L., Jull, A. J. T. & Donahue, D. Accelerator radiocarbon dating at the molecular level. *J. Archaeol. Sci.* **18**, 35–72 (1991).
- Ramsey, C. B. Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).
- Reimer, P. J. *et al.* IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon* **51**, 1111–1150 (2009).
- Yang, D. Y., Eng, B., Wayne, J. S., Dudar, J. C. & Sanders, S. R. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol.* **105**, 539–543 (1998).
- Svensson, E. M. *et al.* Tracing genetic change over time using nuclear SNPs in ancient and modern cattle. *Anim. Genet.* **38**, 378–383 (2007).
- Powell, R. & Gannon, F. Purification of DNA by phenol extraction and ethanol precipitation. Oxford Practical Approach Series. <http://fds.oup.com/www.oup.co.uk/pdf/pas/9v1-7-3.pdf> (2002).
- Orlando, L. *et al.* Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74–78 (2013).
- Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060 (2010).
- Lindgreen, S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res. Notes* **5**, 337 (2012).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Schubert, M. *et al.* Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* **13**, 178 (2012).
- Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43**, 491–498 (2011).
- Krause, J. *et al.* A complete mtDNA genome of an early modern human from Kostenki, Russia. *Curr. Biol.* **20**, 231–236 (2010).
- Skoglund, P., Storå, J., Götherström, A. & Jakobsson, M. Accurate sex identification in ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* **40**, 4477–4482 (2013).
- Rasmussen, M. *et al.* An Aboriginal Australian genome reveals separate human dispersals in Asia. *Science* **334**, 94–98 (2011).
- Frazer, K. A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
- The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- Van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, E386–E394 (2009).

52. Behar, D. M. *et al.* A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
53. Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010).
54. Wei, W. *et al.* A calibrated human Y-chromosomal phylogeny based on resequencing. *Genome Res.* **23**, 388–395 (2013).
55. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
56. Hancock, A. M. *et al.* Adaptations to climate-mediated selective pressures in humans. *PLoS Genet.* **7**, e1001375 (2011).
57. Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
58. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
59. Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
60. Skoglund, P. & Jakobsson, M. Archaic human ancestry in East Asia. *Proc. Natl Acad. Sci. USA* **108**, 18301–18306 (2011).
61. Patterson, N., Price, A. L. & Reich, D. Population structure and Eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
62. Skoglund, P. *et al.* Origins and Genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466–469 (2012).
63. Surakka, I. *et al.* Founder population-specific HapMap panel increases power in GWA studies through improved imputation accuracy and CNV tagging. *Genome Res.* **20**, 1344–1351 (2010).
64. International HapMap3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
65. Busing, F. M. T. A., Meijer, E. & Van der Leeden, R. Delete-m Jackknife for Unequal m. *Stat. Comput.* **9**, 3–8 (1999).
66. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).



# Prefrontal parvalbumin interneurons shape neuronal activity to drive fear expression

Julien Courtin<sup>1,2</sup>, Fabrice Chaudun<sup>1,2</sup>, Robert R. Rozeske<sup>1,2</sup>, Nikolaos Karalis<sup>1,2</sup>, Cecilia Gonzalez-Campo<sup>1,2</sup>, H     Wurtz<sup>1,2</sup>, Azzedine Abdi<sup>3,4</sup>, Jerome Baufreton<sup>3,4</sup>, Thomas C. M. Bienvenu<sup>1,2</sup> & Cyril Herry<sup>1,2</sup>

**Synchronization of spiking activity in neuronal networks is a fundamental process that enables the precise transmission of information to drive behavioural responses<sup>1–3</sup>. In cortical areas, synchronization of principal-neuron spiking activity is an effective mechanism for information coding that is regulated by GABA ( $\gamma$ -aminobutyric acid)-ergic interneurons through the generation of neuronal oscillations<sup>4,5</sup>. Although neuronal synchrony has been demonstrated to be crucial for sensory, motor and cognitive processing<sup>6–8</sup>, it has not been investigated at the level of defined circuits involved in the control of emotional behaviour. Converging evidence indicates that fear behaviour is regulated by the dorsomedial prefrontal cortex<sup>9–12</sup> (dmPFC). This control over fear behaviour relies on the activation of specific prefrontal projections to the basolateral complex of the amygdala (BLA), a structure that encodes associative fear memories<sup>13–15</sup>. However, it remains to be established how the precise temporal control of fear behaviour is achieved at the level of prefrontal circuits. Here we use single-unit recordings and optogenetic manipulations in behaving mice to show that fear expression is causally related to the phasic inhibition of prefrontal parvalbumin interneurons (PVINs). Inhibition of PVIN activity disinhibits prefrontal projection neurons and synchronizes their firing by resetting local theta oscillations, leading to fear expression. Our results identify two complementary neuronal mechanisms mediated by PVINs that precisely coordinate and enhance the neuronal activity of prefrontal projection neurons to drive fear expression.**

To identify the prefrontal circuitry involved in conditioned fear behaviour, mice were implanted with recording electrodes aimed at the dmPFC, and submitted to auditory fear conditioning, a robust learning paradigm in which animals learn to associate a neutral stimulus (the conditioned stimulus, CS) with a coincident aversive foot-shock (the unconditioned stimulus, US) (Fig. 1a). Re-exposure to the CS induces the expression of various conditioned fear responses, including an immobilization reaction called freezing. Twenty-four hours after conditioning, mice displayed a selective increase in freezing during presentations of the CS associated with the US (CS<sup>+</sup>), which returned to baseline levels by the end of the second extinction session (Fig. 1b). One week later, CS<sup>+</sup> presentations induced a selective fear recovery (Fig. 1b). Among the 732 neurons recorded in dmPFC, 493 (67.3%) displayed significant excitatory or inhibitory phasic responses to CS<sup>+</sup> presentations following conditioning. To dissect dmPFC circuits involved in the control of fear behaviour, we separated the CS<sup>+</sup>-responsive neurons into putative principal neurons (PNs,  $n = 351$ ) and interneurons (INs,  $n = 142$ ) using unsupervised clustering and cross-correlogram analyses (Extended Data Fig. 1). Among dmPFC INs, principal component analyses revealed two main subclasses with opposite CS-evoked responses during fear expression (Fig. 1c, d, and Extended Data Fig. 2a, b). Type 1 INs ( $n = 68$ ) displayed short-latency, CS-evoked activity correlated with high (CS<sup>+</sup>), but not low (CS<sup>−</sup>), fear states. Conversely, type 2 INs ( $n = 15$ ) were strongly inhibited during high but not low fear states (Fig. 1c, d). Correlation analyses carried out

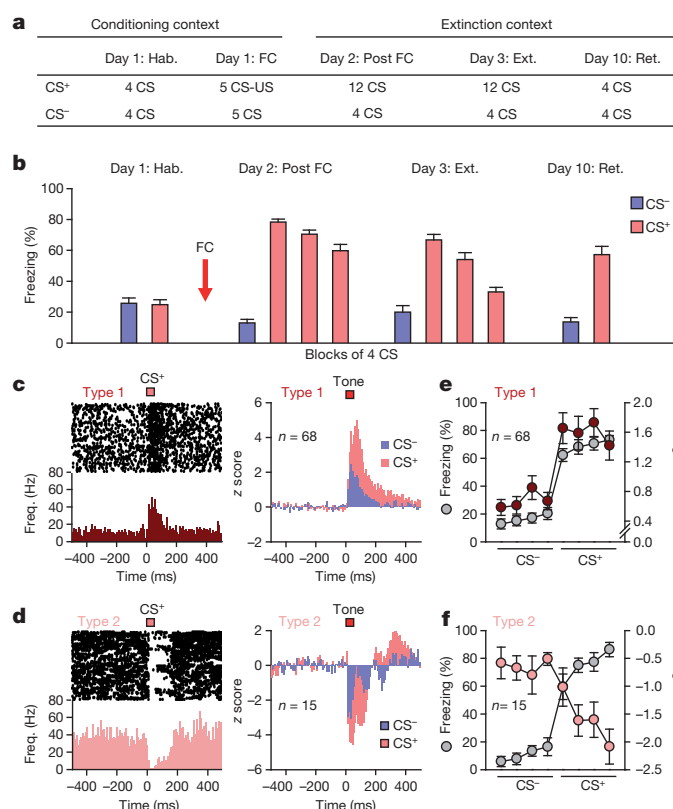
between changes in activity after CS presentations and freezing levels revealed that the firing of type 1 and type 2 INs were correlated and inversely correlated, respectively, with freezing (Fig. 1e, f). Moreover, latency and cross-correlation analyses of simultaneously recorded cells revealed that CS<sup>+</sup>-evoked excitation of type 1 INs preceded type 2 INs CS<sup>+</sup>-evoked inhibition (Extended Data Fig. 2c–e).

Interestingly, whereas type 1 INs displayed moderate firing rates ( $16.2 \pm 1.5$  Hz) and were weakly modulated with local theta oscillations, type 2 INs showed fast firing activity ( $43.9 \pm 9.7$  Hz) and were strongly modulated with local theta, suggesting that type 2 INs are PVINs<sup>16</sup> (Extended Data Fig. 2f–h). To address this possibility, we selectively infected PVINs with injections of a conditional adeno-associated virus (AAV) encoding for archaeorhodopsin in the dmPFC of mice expressing the Cre recombinase under the control of a PV promoter (PV-IRES-Cre; Fig. 2a and Extended Data Fig. 3a, b). Using this strategy, we optically silenced the firing of type 2 ( $n = 5/5$  (5 out of 5)) but not type 1 INs ( $n = 0/9$ ), indicating that type 2 INs belong to the PVIN population (Fig. 2b). Remarkably, among light-reactive PVINs ( $n = 9$ ), only type 2 PVINs ( $n = 5$ ) displayed significant decreases in CS-evoked activity following conditioning, suggesting a functional role of this subpopulation during fear behaviour (Extended Data Fig. 4a–d). In summary, we identified two subclasses of dmPFC INs whose activities oppositely correlate with fear behaviour and demonstrated that type 2 INs are PVINs.

To determine whether the CS-evoked inhibition of type 2 PVINs causes fear expression, PV-IRES-Cre mice received intra-dmPFC injections of a conditional AAV encoding for archaeorhodopsin or channelrhodopsin. Infection of dmPFC PVINs did not change their electrophysiological characteristics (Extended Data Fig. 3c–e). Before fear conditioning, optical silencing of PVINs induced freezing (Fig. 2c). Moreover, after fear extinction, CS<sup>+</sup> presentations coupled to optical silencing of PVINs, including type 2 INs, consistently reinstated fear responses (Fig. 2c and Extended Data Fig. 4c–e). Conversely, optical activation of PVINs transiently inhibited freezing (Fig. 2d). To control that freezing induced by CS-evoked inhibition of type 2 INs did not result from motor impairments, we optically inhibited PVINs during a place avoidance paradigm, in which mice could actively avoid the compartment in which they received optical silencing. Under these conditions, optogenetic silencing of PVINs produced place aversion relative to control animals (Extended Data Fig. 5). These data demonstrate that fear expression is causally related to the inhibition of dmPFC PVINs, including type 2 INs.

PVINs target the perisomatic region of PNs, thereby providing powerful inhibition of dmPFC output activity<sup>17</sup>. Therefore, CS<sup>+</sup>-evoked inhibition of PVINs during fear behaviour might disinhibit PNs, a permissive mechanism that would gate fear responses. Consistent with this, the vast majority of tone-reactive PNs ( $n = 308/351$ , 87.7%) significantly increased their activity upon CS<sup>+</sup> relative to CS<sup>−</sup> presentations (Fig. 3a). Moreover, the optogenetic activation of PVINs inhibited PNs, prevented CS<sup>+</sup>-induced activation of PNs and reduced freezing (Extended Data Fig. 6a–c). Conversely, light-induced inhibition of

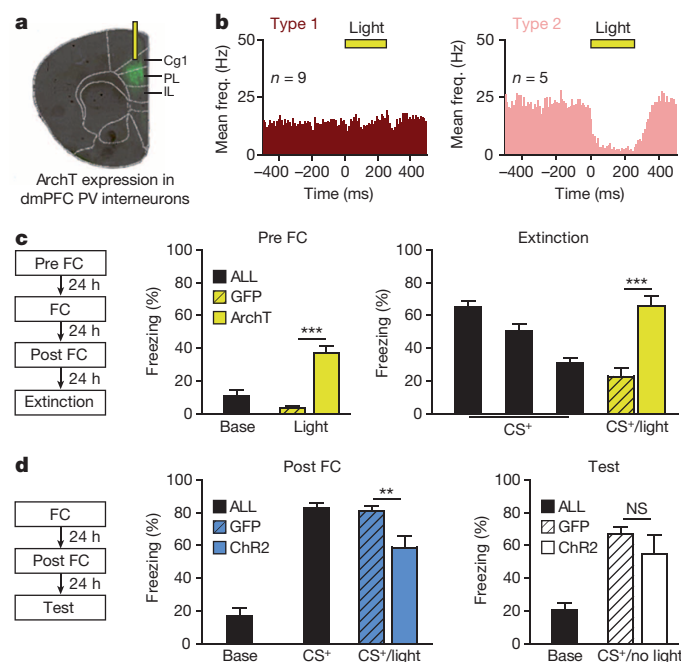
<sup>1</sup>INSERM, Neurocentre Magendie, U862, 146 Rue L    -Saignat, Bordeaux 33077, France. <sup>2</sup>University of Bordeaux, Neurocentre Magendie, U862, 146 Rue L    -Saignat, Bordeaux 33077, France. <sup>3</sup>University of Bordeaux, Institut des Maladies Neurod    g    tives, UMR 5293, Bordeaux F-33000, France. <sup>4</sup>CNRS, Institut des Maladies Neurod    g    tives, UMR 5293, Bordeaux F-33000, France.



**Figure 1 | Firing of distinct dmPFC INs oppositely correlates with fear expression.** **a**, Protocol. **b**, During habituation (Hab.), mice ( $n = 29$ ) exhibited low freezing during CS<sup>-</sup> and CS<sup>+</sup>. After fear conditioning (Post FC; the first extinction session), CS<sup>+</sup> (CS presentations 1–12, grouped into blocks of 4) induced high freezing (Wilcoxon signed-rank tests, CS<sup>-</sup> versus each CS<sup>+</sup> block; all  $P < 0.001$ ). After extinction (Ext. (the second extinction session),  $n = 28$  mice), CS<sup>+</sup> (CS<sup>+</sup> 9–12; the last 4 CS<sup>+</sup> of the extinction session) and CS<sup>-</sup> evoked low freezing. During retrieval (Ret.), CS<sup>+</sup> but not CS<sup>-</sup> induced fear recovery ( $n = 21$  mice, Wilcoxon signed-rank test, CS<sup>-</sup> versus CS<sup>+</sup>;  $P < 0.001$ ). Error bars, mean  $\pm$  s.e.m. **c**, **d**, Left, raster plots and peristimulus time histograms (PSTHs) of CS<sup>+</sup>-evoked firing for INs (type 1 and 2) during Post FC (CS<sup>+</sup> 1–4, 108 trials). Right, mean z score of CS<sup>-</sup> and CS<sup>+</sup>-evoked responses of type 1 and type 2 INs during Post-FC, Ext. or Ret. sessions (CS<sup>-</sup> and CS<sup>+</sup> 1–4, 108 trials). Type 1 INs were excited ( $n = 68$ , 25 mice, paired  $t$ -test, CS<sup>-</sup> versus CS<sup>+</sup>,  $P < 0.001$ ), whereas type 2 INs were inhibited during CS<sup>+</sup> ( $n = 15$ , 8 mice, paired  $t$ -test, CS<sup>-</sup> versus CS<sup>+</sup>,  $P < 0.001$ ). Bins of 10 ms. **e**, **f**, Correlations between freezing during CS (Post-FC, Ext. or Ret. sessions, CS<sup>-</sup> and CS<sup>+</sup> 1–4) and CS-evoked firing (mean z score 0–150 ms post CS) for type 1 INs ( $n = 68$ , Pearson's  $r = 0.79$ ,  $P < 0.01$ ) and type 2 ( $n = 15$ , Pearson's  $r = -0.93$ ,  $P < 0.001$ ).

PVINs disinhibited PNs and produced freezing (Figs 2c and 3b). These data suggest that the increased activity of dmPFC PNs during fear expression results from a disinhibitory mechanism mediated by PVINs.

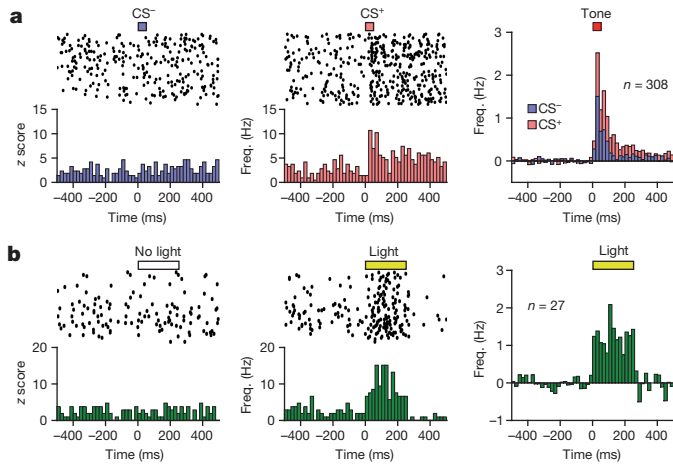
As PVINs have a key role in the genesis of cortical networks oscillations<sup>18,19</sup>, we investigated whether specific changes in dmPFC local field potentials (LFPs) were associated with different fear states. Although freezing periods were associated with a strong reduction of LFP theta power compared to non-freezing periods (Extended Data Fig. 7a), CS<sup>+</sup> but not CS<sup>-</sup> presentations were associated with a transient amplitude increase and a phase resetting of theta oscillations (Fig. 4a and Extended Data Fig. 7b). This analysis produced similar results when restricted to freezing and non-freezing periods during CS presentations (Extended Data Fig. 7c, d). This observation raises the question of whether dmPFC theta phase resetting during fear behaviour is mediated locally or imposed by a remote structure, such as the hippocampus. To address this question we locally injected muscimol to inactivate the medial septum, a structure that is involved in the genesis of hippocampal theta oscillations<sup>20</sup>. Inactivation of the medial septum reduced hippocampal



**Figure 2 | Prefrontal type 2 PVINs control fear expression.** **a**, Schematic of light inhibition of archaerhodopsin (ArchT)-green fluorescent protein (GFP)-expressing PVINs (green) in dmPFC with an optic fibre coupled to the recording electrodes (yellow). Cg1, anterior cingulate cortex; IL, infralimbic area; PL, prelimbic area. **b**, PSTHs showing mean activity changes for type 1 (left,  $n = 9$ ) and type 2 INs (right,  $n = 5$ ) upon yellow light (yellow bars, 250 ms; 108 trials, 0.9 Hz). A Fisher exact statistical test revealed that the proportions of the two populations were significantly different ( $P = 0.033$ ). Bins of 10 ms. **c**, Protocol (left panel) and behaviour from PV-ires-Cre mice infected in dmPFC with GFP (control,  $n = 8$ ) or ArchT-GFP-expressing ( $n = 9$ ) floxed AAV viruses and submitted to yellow light. Before conditioning (Pre FC, middle panel) and after extinction (right panel), optogenetic inhibition of PVINs induced freezing (paired  $t$ -tests, Pre FC, GFP versus ArchT,  $***P < 0.001$ ; Extinction, GFP versus ArchT,  $***P < 0.001$ ; light-pulse duration, 250 ms; 108 trials, 0.9 Hz). **d**, Protocol (left panel) and behaviour from PV-ires-Cre mice infected with control GFP ( $n = 8$ ) or channelrhodopsin (ChR2)-enhanced yellow fluorescent protein (eYFP)-expressing ( $n = 6$ ) floxed AAV viruses in the dmPFC and submitted to blue light. Following conditioning (middle panel, Post FC), optogenetic activation of PVINs decreased freezing (Post FC, GFP versus ChR2, paired  $t$ -test,  $**P < 0.01$ ; light-pulse duration, 250 ms; 108 trials, 0.9 Hz). NS, not significant. Error bars, mean  $\pm$  s.e.m.

theta power, whereas it did not influence freezing and had no effect on dmPFC theta phase resetting evoked by CS<sup>+</sup> presentations (Extended Data Fig. 8).

Interestingly, we observed a strong correlation between CS<sup>+</sup>-evoked inhibition of PVINs and dmPFC theta phase resetting, suggesting that this phenomenon is gated by PVINs (Fig. 4b). In support of this hypothesis, optogenetic inhibition of PVINs reproduced theta phase resetting (Fig. 4c and Extended Data Fig. 9). Consistent with this, dmPFC theta resetting induced by CS<sup>+</sup> was blocked by optogenetic excitation of PVINs (Fig. 4d). Our results indicate that CS<sup>+</sup>-evoked inhibition of PVINs mediates theta phase resetting during fear expression, a phenomenon that might enhance synchronization and efficiency of dmPFC output neurons. To evaluate whether dmPFC theta phase resetting is associated with spiking synchronization among PNs during fear expression, we quantified the number of PNs displaying a significant firing increase during CS<sup>-</sup> and CS<sup>+</sup> presentations. Significantly more PNs were activated during CS<sup>+</sup> relative to CS<sup>-</sup> presentations (Fig. 5a). This activation was associated with a significant increase of coincident firing between pairs of PNs following CS<sup>+</sup> (Fig. 5b and Extended Data Fig. 10a). Furthermore, more PNs were significantly phase-locked to local theta oscillations during CS<sup>+</sup> relative to CS<sup>-</sup> presentations (Fig. 5c). Consistent with this, comparison of



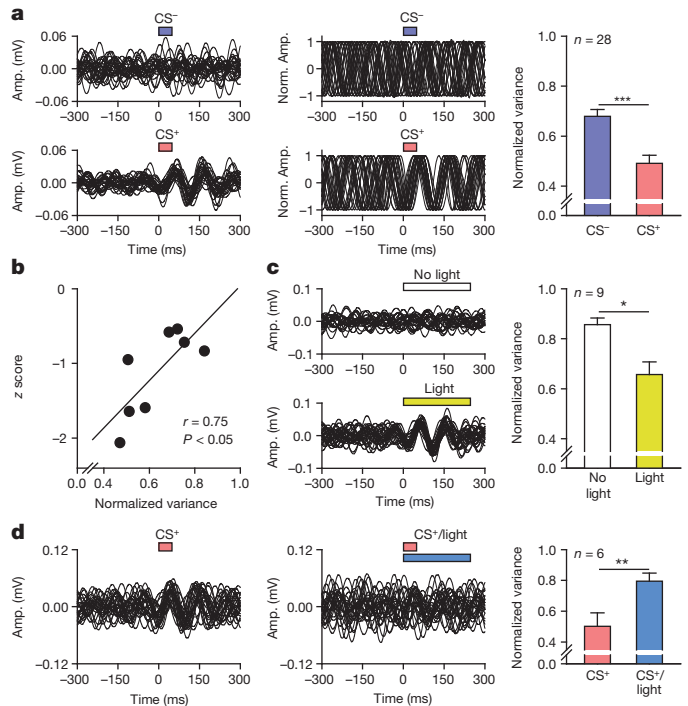
**Figure 3 | dmPFC PVINs disinhibit PNs during fear expression.** **a**, Raster plots and PSTHs illustrating CS<sup>-</sup> and CS<sup>+</sup>-evoked activities (left and middle panels, respectively) of a PN during Post FC (CS<sup>-</sup>/CS<sup>+</sup> 1–4, 108 trials). Right panel, mean PSTHs of PNs recorded during Post FC, Ext. or Ret. ( $n = 308$  neurons from 27 mice, CS<sup>-</sup> and CS<sup>+</sup> 1–4) showing a stronger and significant increase in response to CS<sup>+</sup> compared to CS<sup>-</sup>. **b**, Firing of a PN recorded in a mouse expressing ArchT in dmPFC PVINs at baseline (left panel, no light), and in response to yellow light (middle panel, light-pulse duration, 250 ms; 108 trials, 0.9 Hz). Right panel, mean PSTHs of all PNs displaying significant CS<sup>+</sup>-evoked excitation during Ext. and disinhibited during optogenetic inhibition of PVINs ( $n = 27/41$  PNs from 7 mice; light-pulse duration, 250 ms; 108 trials, 0.9 Hz). Bins of 20 ms.

the strength of theta phase locking, a measure of spiking synchronization, revealed a stronger tuning of dmPFC activity to local theta during CS<sup>+</sup> periods (Extended Data Fig. 10b). To evaluate whether enhancement of the spiking synchronization of PNs with local theta induced by CS<sup>+</sup> presentations was causally related to the inhibition of PVINs, we optogenetically manipulated PVINs and quantified PN theta phase locking. Our analysis revealed that light-induced inhibition of PVINs increased, whereas light-induced excitation of PVINs reduced PNs phase locking to dmPFC theta oscillations (Extended Data Fig. 10c, d).

To understand the dynamics of PNs synchronization during theta phase reset, the mean preferred phase of individual PNs was calculated during the first three theta cycles following CS<sup>+</sup> (Supplementary Methods). Relative to CS<sup>-</sup> presentations, CS<sup>+</sup>-induced firing of PNs occurred significantly more frequently around the peak of the oscillations, thereby creating precise temporal windows during which PNs were synchronized (Fig. 5d). Interestingly, similar to the CS<sup>+</sup> condition, artificial resetting of local theta oscillations, either by aligning the phase of individual LFPs during CS<sup>-</sup> presentations or by optogenetically inhibiting PVINs, produced synchronization of PNs firing around the peak of theta oscillations (Extended Data Fig. 10e, f). This observation suggests that the overall phase preference of PNs did not change between CS<sup>-</sup> or CS<sup>+</sup> conditions, but that PV-mediated theta phase resetting coordinated and sharpened synchronization among PNs.

Converging evidence indicates that dmPFC PNs target both the basolateral amygdala (BLA) and the periaqueductal grey (PAG), two structures involved in fear behaviour<sup>21–23</sup>. This raises the possibility that PNs may modulate fear expression through direct projections to the PAG and/or the BLA. To disentangle these possibilities, we antidromically activated dmPFC efferents using extracellular stimulation of BLA or PAG in anaesthetized mice, following completion of behaviour. These experiments revealed that PNs disinhibited during CS<sup>+</sup> presentations preferentially targeted the BLA (Fig. 5e, f and Extended Data Fig. 6d). These data indicate that theta phase resetting mediated by PVINs synchronizes PNs after CS<sup>+</sup> presentations and suggest that dmPFC PNs preferentially target the BLA to drive fear responses.

Using single-unit and LFP recordings combined with optogenetic manipulation of PVINs in mice, we have shown that a subpopulation

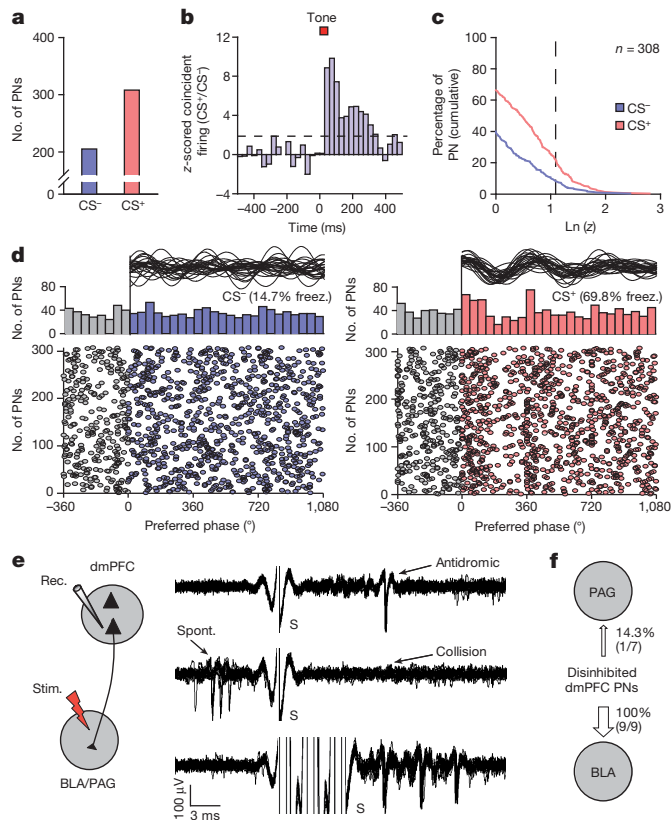


**Figure 4 | Inhibition of PVINs induces theta phase resetting.** **a**, dmPFC LFP trace amplitudes (Amp.) filtered in the 8–12-Hz range (left panel, 27 sweeps) and corresponding standardized trace amplitudes (middle panel) illustrating theta phase resetting induced by CS<sup>+</sup> but not CS<sup>-</sup> (Post-FC session, first CS<sup>-</sup> and first CS<sup>+</sup>, 27 sweeps). Right panel, time variance of the first theta peak following CS (Post FC, CS<sup>-</sup> and CS<sup>+</sup> 1–4,  $n = 28$  mice, CS<sup>-</sup> versus CS<sup>+</sup>, paired  $t$ -test, \*\*\* $P < 0.001$ ). **b**, Correlation between CS-evoked firing of type 2 INs (mean  $z$  score 0–150 ms post CS) and the time variance of the first theta peak following CS (Post FC, CS<sup>-</sup> and CS<sup>+</sup> 1–4,  $n = 15$  type 2 INs, Pearson's  $r = 0.75$ ,  $P < 0.05$ ). **c**, Left panel, dmPFC LFP traces recorded in a mouse expressing ArchT in PVINs in control conditions (top part; No light, 27 sweeps), and during optogenetic inhibition of PVINs (bottom part; light duration, 250 ms; 27 sweeps, 0.9 Hz). Right panel, time variance of the first theta peak following light ( $n = 9$  mice, No light versus Light, paired  $t$ -test, \* $P < 0.05$ ). **d**, Representative LFP traces recorded when CS<sup>+</sup> was presented alone (left panel, Post-FC session, first CS<sup>+</sup>) or paired with the optogenetic activation of PVINs (middle panel, Post-FC session, fifth CS<sup>+</sup>; light-pulse duration, 250 ms; 27 sweeps, 0.9 Hz). Right panel, time variance analyses of theta resetting during CS<sup>+</sup> or CS<sup>+</sup>/Light ( $n = 6$  mice, CS<sup>+</sup> versus CS<sup>+</sup>/Light, paired  $t$ -test, \*\* $P < 0.01$ ). Error bars, mean  $\pm$  s.e.m.

of PVINs organizes the spiking activity of dmPFC PNs during precise time windows, through phase resetting of local theta oscillations, to drive fear expression. Our data indicate that the fine regulation of dmPFC-BLA PNs by a subtype of PVINs is critical for the expression of fear behaviour. Our results demonstrate that inhibition of type 2 PVINs during CS<sup>+</sup> presentations is causally related to the expression of conditioned fear responses, and suggest that type 1 INs might inhibit type 2 INs. The origin of CS-mediated type 1 INs excitatory responses remains to be determined, but it is likely that they receive inputs from structures involved in the encoding or modulation of conditioned fear such as BLA or hippocampus<sup>24,25</sup>.

A key question is what mechanisms can account for our observation that inhibition of PVINs is necessary and sufficient for the expression of fear responses. Cortical PVINs are known to inhibit PNs through powerful perisomatic inhibition<sup>17</sup>. As a consequence, CS<sup>+</sup>-evoked inhibition in PVINs induced a strong disinhibition of PNs, a permissive mechanism that gated neuronal responses during fear expression. These results indicate that CS-evoked activity in dmPFC PNs during fear expression result in part from a disinhibitory mechanism. Notably, conditioned freezing was not entirely prevented by PVINs activation, indicating that some dmPFC PNs may escape inhibitory control, or





**Figure 5 | Synchronization of dmPFC PNs during fear expression.**

**a**, Number of significantly CS-activated PNs recorded during Post-FC, Ext. or Ret. sessions ( $CS^-$ ,  $n = 205$ ;  $CS^+$  1–4,  $n = 308$ ;  $CS^-$  versus  $CS^+$ , paired  $t$ -test,  $P < 0.001$ ). **b**, Normalized averaged ratio of changes in coincident activity between pairs of PNs induced by  $CS^-$  and  $CS^+$  (Post-FC, Ext. or Ret. sessions,  $n = 975$  pairs from 308 PNs). Dashed line, significant  $z$  score at  $P < 0.05$  level. Bins of 30 ms. **c**, Cumulative distribution of log-transformed Rayleigh's test  $Z$  of CS-responsive PNs ( $n = 308$ ; Post-FC, Ext. or Ret. sessions). Dashed line, significant theta phase locking threshold ( $\ln(Z) = 1.1$ ,  $P < 0.05$ ;  $CS^-$ ,  $n = 24$  neurons;  $CS^+$  1–4,  $n = 65$  neurons). **d**, Top panel, preferred theta phase distributions of PNs ( $n = 308$  neurons, 108 CS pips, Post-FC, Ext. or Ret. sessions) during theta cycles around  $CS^-$  (left part, blue bars, 14.7% freezing) and  $CS^+$  (right part, red bars, 69.8% freezing, bins of  $45^\circ$ ). Bottom panel, preferred theta phases of individual PNs. Example 8–12-Hz theta-filtered LFP traces during  $CS^-$  and  $CS^+$  are represented above for illustrative purposes. During  $CS^+$  but not  $CS^-$ , resetting of theta oscillations synchronizes the firing of PNs around the peaks of theta cycles (Rayleigh test for circular uniformity: first theta cycle post CS,  $CS^+$  versus  $CS^-$ ,  $P < 0.001$ ). **e**, Left panel, strategy used to identify connections between PNs and the BLA–PAG. Rec., recording electrode; S, stimulation artifact; Stim., stimulation electrode. Right panel, antidromic spikes recorded from a PN in response to BLA stimulations identified by their low temporal jitter (top trace, 10 trials), collisions with spontaneously (Spont.) occurring spikes (middle trace, 10 trials) and ability to follow high-frequency stimulation (bottom trace, 250 Hz, 10 trials). **f**, PNs exhibiting antidromic responses to BLA stimulations displayed  $CS^+$ -evoked excitation (9/9 neurons). Only a small fraction of PNs exhibiting antidromic responses to PAG stimulation displayed  $CS^+$ -evoked excitation (1/7 neurons, 14.3%). Thin arrow indicates that fewer neurons that project to the PAG are disinhibited; thick arrow indicates that more neurons that project to the BLA are disinhibited.

that other brain regions promote fear responses in concert with dmPFC.

Although fear behaviour was associated with a reduction in dmPFC theta-oscillation magnitude,  $CS^+$ -evoked inhibition of PVINs induced a robust and transient theta phase resetting spanning two to three theta cycles. Transient theta phase has been previously observed in cortical regions following electrical or sensory stimulations<sup>26–28</sup>. Our findings provide the first mechanistic explanation of phase resetting at the

cellular level and extend this phenomenon to the control of emotional behaviour. Functionally, we observed that theta phase resetting synchronized PNs around theta peaks without changing the preferred phases of individual PNs. This observation suggests that resetting of local theta oscillations, but not the preferred phases of individual PNs to the local LFP, is critically involved in the expression of fear responses. Thus, theta phase resetting represents a powerful mechanism for reliable fear expression because it creates an optimal temporal relationship that binds spiking activity with sensory information provided by CS. Ultimately, phase resetting of oscillations is a powerful mechanism that enhances the impact of input signals and enables transmission of information to downstream targets. Our data also show that reduction of rhythmic inhibition from PVINs paradoxically increases synchrony. Suppression of interference between two oscillators may account for this effect. Future work will be needed to identify the origin of dmPFC theta oscillations that are unmasked by PVINs inhibition.

Another question is how synchronized PNs can control fear expression. Previous findings suggest that putative dmPFC PNs displaying sustained or transient changes in their spiking activity promote fear expression through activation and synchronization of BLA neurons<sup>9,29,30</sup>. In line with these studies, our results demonstrate that PNs exhibiting  $CS^+$ -evoked synchronized firing during fear expression preferentially project to the BLA where they may target specific neuronal populations activated during fear behaviour<sup>23</sup>.

Finally, our findings suggest that persistent fear behaviour, which is at the core of psychiatric conditions such as anxiety disorders, may be finely regulated at the level of specific prefrontal inhibitory circuits.

## METHODS SUMMARY

Mice were submitted to a fear-conditioning paradigm in which the  $CS^+$  but not the  $CS^-$  was paired with a mild foot-shock (US). Extinction training was carried out over 2 days and mice were tested 1 week later for a retrieval session<sup>23</sup>. For optogenetic manipulations, PV-IRES-Cre mice received stereotaxic injections of AAV viruses encoding channelrhodopsin or archaeorhodopsin in the dmPFC. Bilateral activation of archaeorhodopsin or channelrhodopsin was performed using implanted optic fibres coupled to a laser beam. Inactivation of the medial septum was achieved using local pressure injection of fluorescently labelled muscimol. Individual neurons were recorded extracellularly and spikes were sorted by time-amplitude window discrimination and template matching as described<sup>23</sup>. CS-evoked responses were normalized to baseline activity using a  $z$ -score transformation. Antidromic and orthodromic spikes evoked by extracellular stimulations of the BLA or PAG were recorded in neurons isolated from behavioural sessions and recorded in urethane-anaesthetized mice, after completion of behaviour. *In vitro* whole-cell voltage and current-clamp recordings were performed using glass pipettes (4–6 M $\Omega$ ) filled with K-gluconate-based solutions.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 15 May; accepted 9 October 2013.**

**Published online 20 November 2013.**

- Singer, W. Neuronal synchrony: a versatile code for the definition of relations? *Neuron* **24**, 49–65, 111–125 (1999).
- Buzsáki, G. & Draguhn, A. Neuronal oscillations in cortical networks. *Science* **304**, 1926–1929 (2004).
- Womelsdorf, T. *et al.* Modulation of neuronal interactions through neuronal synchronization. *Science* **316**, 1609–1612 (2007).
- Royer, S. *et al.* Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition. *Nature Neurosci.* **15**, 769–775 (2012).
- Cobb, S. R., Buhl, E. H., Halasy, K., Paulsen, O. & Somogyi, P. Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons. *Nature* **378**, 75–78 (1995).
- Benchenane, K. *et al.* Coherent theta oscillations and reorganization of spike timing in the hippocampal–prefrontal network upon learning. *Neuron* **66**, 921–936 (2010).
- Friedrich, R. W., Habermann, C. J. & Laurent, G. Multiplexing using synchrony in the zebrafish olfactory bulb. *Nature Neurosci.* **7**, 862–871 (2004).
- Riehle, A., Grun, S., Diesmann, M. & Aertens, A. Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* **278**, 1950–1953 (1997).



9. Burgos-Robles, A., Vidal-Gonzalez, I. & Quirk, G. J. Sustained conditioned responses in prelimbic prefrontal neurons are correlated with fear expression and extinction failure. *J. Neurosci.* **29**, 8474–8482 (2009).
10. Tang, J. *et al.* Pavlovian fear memory induced by activation in the anterior cingulate cortex. *Mol. Pain* **1**, 6 (2005).
11. Vidal-Gonzalez, I., Vidal-Gonzalez, B., Rauch, S. L. & Quirk, G. J. Microstimulation reveals opposing influences of prelimbic and infralimbic cortex on the expression of conditioned fear. *Learn. Mem.* **13**, 728–733 (2006).
12. Corcoran, K. A. & Quirk, G. J. Activity in prelimbic cortex is necessary for the expression of learned, but not innate, fears. *J. Neurosci.* **27**, 840–844 (2007).
13. Pape, H. C. & Pare, D. Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear. *Physiol. Rev.* **90**, 419–463 (2010).
14. Knapska, E. *et al.* Functional anatomy of neural circuits regulating fear and extinction. *Proc. Natl Acad. Sci. USA* **109**, 17093–17098 (2012).
15. LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
16. Hartwich, K., Pollak, T. & Klausberger, T. Distinct firing patterns of identified basket and dendrite-targeting interneurons in the prefrontal cortex during hippocampal theta and local spindle oscillations. *J. Neurosci.* **29**, 9563–9574 (2009).
17. Freund, T. F. & Katona, I. Perisomatic inhibition. *Neuron* **56**, 33–42 (2007).
18. Blatow, M. *et al.* A novel network of multipolar bursting interneurons generates theta frequency oscillations in neocortex. *Neuron* **38**, 805–817 (2003).
19. Losonczy, A., Zemelman, B. V., Vaziri, A. & Magee, J. C. Network mechanisms of theta related neuronal activity in hippocampal CA1 pyramidal neurons. *Nature Neurosci.* **13**, 967–972 (2010).
20. Yoder, R. M. & Pang, K. C. Involvement of GABAergic and cholinergic medial septal neurons in hippocampal theta rhythm. *Hippocampus* **15**, 381–392 (2005).
21. Gabbott, P. L., Warner, T. A., Jays, P. R., Salway, P. & Busby, S. J. Prefrontal cortex in the rat: projections to subcortical autonomic, motor, and limbic centers. *J. Comp. Neurol.* **492**, 145–177 (2005).
22. Di Scala, G., Mana, M. J., Jacobs, W. J. & Phillips, A. G. Evidence of Pavlovian conditioned fear following electrical stimulation of the periaqueductal grey in the rat. *Physiol. Behav.* **40**, 55–63 (1987).
23. Herry, C. *et al.* Switching on and off fear by distinct neuronal circuits. *Nature* **454**, 600–606 (2008).
24. Sotres-Bayon, F., Sierra-Mercado, D., Pardilla-Delgado, E. & Quirk, G. J. Gating of fear in prelimbic cortex by hippocampal and amygdala inputs. *Neuron* **76**, 804–812 (2012).
25. Tierney, P. L., Degenetais, E., Thierry, A. M., Glowinski, J. & Gioanni, Y. Influence of the hippocampus on interneurons of the rat prefrontal cortex. *Eur. J. Neurosci.* **20**, 514–524 (2004).
26. McCartney, H., Johnson, A. D., Weil, Z. M. & Givens, B. Theta reset produces optimal conditions for long-term potentiation. *Hippocampus* **14**, 684–687 (2004).
27. Buzsáki, G., Grastyán, E., Tveritskaya, I. N. & Czopf, J. Hippocampal evoked potentials and EEG changes during classical conditioning in the rat. *Electroencephalogr. Clin. Neurophysiol.* **47**, 64–74 (1979).
28. Rizzuto, D. S. *et al.* Reset of human neocortical oscillations during a working memory task. *Proc. Natl Acad. Sci. USA* **100**, 7931–7936 (2003).
29. Livneh, U. & Paz, R. Amygdala-prefrontal synchronization underlies resistance to extinction of aversive memories. *Neuron* **75**, 133–142 (2012).
30. Chang, C. H., Berke, J. D. & Maren, S. Single-unit activity in the medial prefrontal cortex during immediate and delayed extinction of fear in rats. *PLoS ONE* **5**, e11971 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank members of the Herry laboratory, K. Benchenane and D. Dupret for comments on the manuscript, K. Deisseroth and E. Boyden for generously sharing material, J. Babelo, S. Wolff and P. Tovote for technical and computational assistance, the Bordeaux Imaging center of the University of Bordeaux, and C. Poujol and S. Marais for technical assistance with microscopy. This work was supported by grants from the French National Research Agency (ANR-2010-BLAN-1442-01; ANR-10-EQPX-08 OPTOPATH), the European Research Council (ERC) under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC grant agreement no. 281168, a Fonds AXA pour la recherche doctoral fellowship (J.C.) and the Conseil Régional d'Aquitaine. T.C.M.B. is a fellow of Ecole de l'Inserm Liliane Bettencourt-MD-PhD program, France.

**Author Contributions** J.C., F.C., R.R.R., N.K., C.G.-C., H.W., A.A., J.B. and T.C.M.B. performed the experiments and analysed the data. J.C. and C.H. designed the experiments and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.H. ([cyril.herry@inserm.fr](mailto:cyril.herry@inserm.fr)).

## METHODS

**Animals.** Male C57BL6/J mice (3 months old, Janvier) and PV-IRES-Cre mice (3 months old, Jackson Laboratory, B6;129P2-Pvalb<sup>tm1(Cre)Arbr/J</sup>) were individually housed for at least 7 days before all experiments, under a 12-h light–dark cycle, and provided with food and water *ad libitum*. All procedures were performed in accordance with standard ethical guidelines (European Communities Directive 86/60-EEC) and were approved by the committee on Animal Health and Care of Institut National de la Santé et de la Recherche Médicale and French Ministry of Agriculture and Forestry (authorization A3312001).

**Behaviour.** Fear conditioning and extinction took place in two different contexts (context A and B). The conditioning and extinction boxes were cleaned with 70% ethanol and 1% acetic acid before and after each session, respectively. To score freezing behaviour, an automated infrared beam detection system located on the bottom of the experimental chambers was used (Coulbourn Instruments). The animals were considered to be freezing if no movement was detected for 2 s. On day 1, C57BL6/J mice were submitted to an habituation session in context A, in which they received four presentations of the CS<sup>+</sup> and of the CS<sup>−</sup> (total CS duration, 30 s; consisting of 50-ms pips at 0.9 Hz repeated 27 times, 2 ms rise and fall; pip frequency, 7.5 kHz or white-noise, 80 dB sound pressure level). Discriminative fear conditioning was performed on the same day by pairing the CS<sup>+</sup> with a US (1-s foot-shock, 0.6 mA, 5 CS<sup>+</sup>–US pairings; inter-trial intervals, 20–180 s). The onset of the US coincided with the offset of the CS<sup>+</sup>. The CS<sup>−</sup> was presented after each CS<sup>+</sup>–US association but was never reinforced (five CS<sup>−</sup> presentations; inter-trial intervals, 20–180 s). The frequencies used for CS<sup>+</sup> and CS<sup>−</sup> were counterbalanced across animals. On day 2 and day 3, conditioned mice were submitted to extinction training (post-fear-conditioning and extinction sessions) in context B during which they received 4 and 12 presentations of the CS<sup>−</sup> and CS<sup>+</sup>, respectively. Retrieval of fear was tested 7 days later in context B, with 4 presentations of the CS<sup>−</sup> and the CS<sup>+</sup>. Four distinct behavioural experiments were performed to collect the entire data set.

For optogenetic experiments using archaeorhodopsin, PV-IRES-Cre mice were submitted on day 1 to a pre-fear-conditioning session in context A during which they received yellow light stimulations (250-ms pulses repeated at 0.9 Hz during 2 min). Fear conditioning was performed on day 2 in context A, by pairing the CS<sup>+</sup> with the US (1-s foot-shock, 0.6 mA, 5 CS<sup>+</sup>/US pairings; inter-trial interval, 20–180 s). On day 2 and day 3, conditioned mice were submitted to extinction training (post-fear-conditioning and extinction sessions) in context B during which they received 12 presentations of the CS<sup>+</sup>. At the end of the last extinction session they received an additional four presentations of the CS<sup>+</sup> coupled to yellow light stimulations (each CS<sup>+</sup> pip was paired with a 250-ms light pulse). For optogenetic experiments using channelrhodopsin, two distinct behavioural experiments were performed to collect the entire data set. For optogenetic experiments using channelrhodopsin, PV-IRES-Cre mice were submitted on day 1 to the same fear conditioning protocol as above. A post-conditioning test was performed on day 2 in context B and consisted of four presentations of the CS<sup>+</sup> alone followed by four presentations of the CS<sup>+</sup> coupled to blue light stimulations (each CS<sup>+</sup> pip was paired with a 250-ms light pulse). On day 3, mice were submitted to a second test in context B (Test) in which they received four presentations of the CS<sup>+</sup>. For optogenetic experiments using channelrhodopsin, two distinct behavioural experiments were performed to collect the entire data set.

For the place-avoidance experiment, we used an apparatus composed of two plexiglas compartments (20 × 10 cm each) connected by an alleyway. The two compartments differed tactilely (smooth plastic versus metal bars) and visually (grey plexiglas with red horizontal stripes or grey plexiglas). The time spent in each compartment was automatically recorded by an infrared beam detection system located on the bottom of the apparatus (Imetronic). On day 1, mice were allowed to explore freely the entire apparatus during a 15-min pre-exposure. Following pre-exposure, the compartment in which the mice spent the most time was designated as the most visited compartment. On day 2, mice were submitted to a 15-min test session during which light pulses (250-ms pulse width, repeated at 0.9 Hz) were delivered while the animals occupied the most visited compartment, but not when they occupied the less-visited compartment. The automated infrared beam sensors detected when the animal fully entered and exited the most visited compartment on day 1. The laser was automatically turned on for the period of time in which the animal stayed in the most visited compartment. The laser was turned on only when the animal fully entered the most visited compartment, not before the entrance. For place avoidance experiments, two distinct behavioural experiments were performed to collect the entire data set.

For pharmacological experiments, C57BL6/J mice were submitted to a fear conditioning paradigm consisting of CS<sup>+</sup> and US pairings in context A as described above. On days 2, 3 and 4, conditioned mice were tested in context B during which they received four presentations of the CS<sup>+</sup> before muscimol injections (Day 2, Test pre-MUS), 5 min after muscimol injections (Day 3, Test MUS), and 24 hrs

following muscimol injections (Day 4, Test post-MUS). For pharmacological experiments, two distinct behavioural experiments were performed to collect the entire data set.

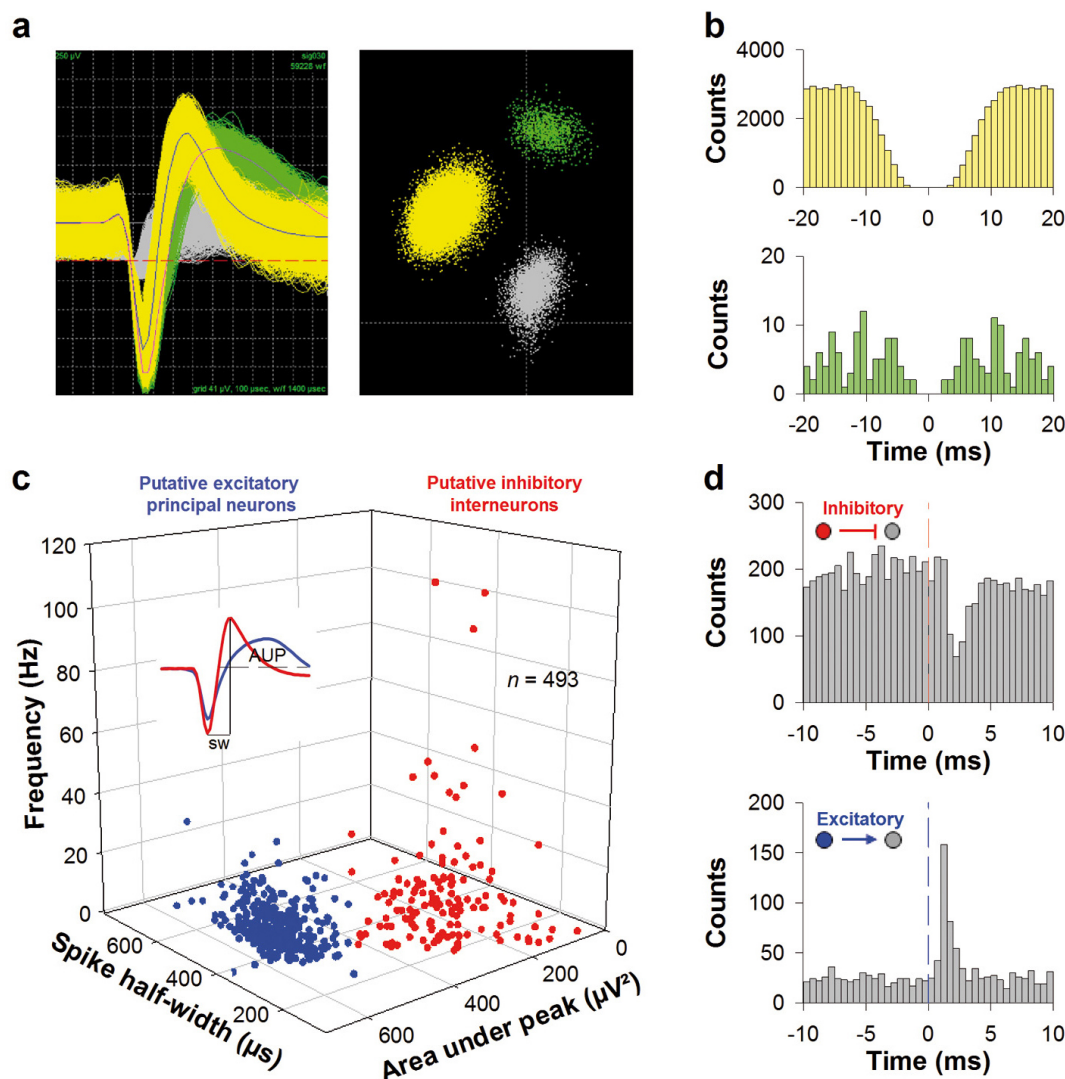
**Surgery and recordings.** Mice were anaesthetized with isoflurane (induction 3%, maintenance 1.5%) in O<sub>2</sub>. Body temperature was maintained at 37 °C with a temperature controller system (FHC). Mice were secured in a stereotaxic frame and unilaterally implanted in the left dorsomedial prefrontal cortex (dmPFC) with a multi-wire electrode array aimed at the following coordinates: 2 mm anterior to bregma; 0.3 mm lateral to the midline; and 0.8 to 1.4 mm ventral to the cortical surface. A subset of animals (*n* = 10) were also implanted in the dorsal hippocampus (dHip) at the following coordinates: 2 mm posterior to bregma; 1.2 mm lateral to midline; and 1.2 to 1.4 mm ventral to the cortical surface. The electrodes consisted of 16 individually insulated nichrome wires (13 µm inner diameter, impedance 30–100 KΩ; Kanthal) contained in a 26-gauge stainless-steel guide cannula. The wires were attached to an 18-pin connector (Omnetics). For mice that received dmPFC and dHip multi-wire implants, two connectors were used. All implants were secured using Super-Bond cement (Sun Medical). After surgery mice were allowed to recover for 7 days and were habituated to handling. Analgesia was applied before, and 1 day after surgery (Metacam, Boehringer). Electrodes were connected to a headstage (Plexon) containing sixteen unity-gain operational amplifiers. The headstage was connected to a 16-channel preamplifier (gain 100× bandpass filter from 150 Hz to 9 kHz for unit activity and from 0.7 Hz to 170 Hz for field potentials, Plexon). Spiking activity was digitized at 40 kHz and bandpass filtered from 250 Hz to 8 kHz, and isolated by time-amplitude window discrimination and template matching using a Multichannel Acquisition Processor system (Plexon). At the conclusion of the experiment, recording sites were marked with electrolytic lesions before perfusion, and electrode tips locations were reconstructed with standard histological techniques.

**Single-unit analyses.** Single-unit spike sorting was performed using Off-Line Spike Sorter (OFSS, Plexon) for all behavioural sessions. Principal-component scores were calculated for unsorted waveforms and plotted in a three-dimensional principal-component space; clusters containing similar valid waveforms were manually defined. A group of waveforms were considered to be generated from a single neuron if the waveforms formed a discrete, isolated, cluster in the principal-component space and did not contain a refractory period less than 1 ms, as assessed using auto-correlogram analyses. To avoid analysis of the same neuron recorded on different channels, we computed cross-correlation histograms. If a target neuron presented a peak of activity at a time that the reference neuron fired, only one of the two neurons was considered for further analysis. After fear conditioning, if the same neuron was sequentially recorded during different behavioural sessions, we considered only the first behavioural session in which it was recorded. To separate putative inhibitory interneurons (INs) from putative excitatory principal neurons (PNs) we used an unsupervised cluster algorithm based on Ward's method. In brief, the Euclidian distance was calculated between all neuron pairs based on the three-dimensional space defined by each neuron's average half-spike width (measured from trough to peak), the firing rate and the area under the hyperpolarization phase of the spike. An iterative agglomerative procedure was then used to combine neurons into groups based on the matrix of distances such that the total number of groups was reduced to give the smallest possible increase within-group sum of square deviation. To assess the significance of cross-correlogram analyses performed between pairs of recorded neurons, a mean firing rate with 95% confidence limits of the target neuron was calculated. Significant short-latency inhibitory or excitatory interactions were retained if the number of action potentials of the target neuron was inferior or superior to the 95% confidence limits, respectively. Moreover, to show that cross-correlations were not simply occurring by chance or were due to CS presentations, we performed two controls. First, the spike train of the neuron was shuffled 100 times and a shuffled cross-correlogram was computed. Absence of short-latency interaction in the shuffled cross-correlogram was indicative that the cross-correlations were not due to chance. Second, to control that short-latency interactions were not artificially induced by stimulus presentations, we computed a shift predictor and subtracted it from the original cross-correlogram. Persistence of short-latency cross-correlations indicates that the neuronal interactions were not due to CS presentations. CS or light-induced neural activity of recorded neurons was calculated by comparing the firing rate after stimulus onset with the firing rate recorded during the 500 ms before stimulus onset (bin size of 10 ms) using a *z*-score transformation. *z*-score values were calculated by subtracting the average baseline firing rate established over the 500 ms preceding stimulus onset from individual raw values and by dividing the difference by the baseline standard deviation. Only CS<sup>+</sup> responsive neurons (at least one significant positive or negative *z*-score bin (*z*-score > ± 1.67, *P* < 0.05) within 100 ms following CS onset) were considered for further analysis. For statistical analysis, *z*-score comparisons were performed using the average *z*-score value calculated during the 150 ms after CS onset.

To identify the main firing patterns among INs, we used an unbiased principal component analysis (PCA) based on the neuronal activity evoked by CS<sup>+</sup> presentations (z-score 500 ms before and after CS<sup>+</sup> presentations, CS<sup>+</sup> presentations 1–4 in post-fear-conditioning, extinction and retrieval sessions, each CS<sup>+</sup> consisting of 27 individual sound pips; bin size of 10 ms). Only the first principal component was considered (PC1) because it explained most of the variance of our data set. Type 1 and type 2 interneurons were defined as correlated and inversely correlated, respectively, with PC1 at the  $P < 0.001$  significance level. Co-firing between recorded PNs pairs was established by quantifying the number of non-overlapping 30-ms time windows following CS<sup>+</sup> presentations during which co-firing events occurred (each pip presentation, CS<sup>+</sup> presentations 1–4, 108 pips, post-fear-conditioning, extinction or retrieval sessions). We then calculated a ratio of coincident firing by dividing the number of co-firing occurrences during CS<sup>+</sup> presentations by those obtained during CS<sup>−</sup> presentations. This coincident firing ratio was normalized to the pre-CS period (500 ms pre CS) using a z-score transformation. To control that the changes in coincident firing between CS<sup>+</sup> and CS<sup>−</sup> conditions were not due to an increase in PNs firing rate during CS<sup>+</sup> presentations, the same analysis was performed but this time the number of co-firing events in each 30-ms time window was normalized by the total number of spikes of the two neurons in this particular time window. Statistical analyses were performed using paired Student's *t*-tests post hoc comparisons at the  $P < 0.05$  level of significance unless indicated otherwise. Results are presented as mean  $\pm$  s.e.m.

**Statistical analyses.** For each statistical analysis provided in the manuscript, the Kolmogorov–Smirnov normality test was first performed on the data to determine whether parametric or non-parametric tests were required. Two different approaches were used to calculate the sample size. For studies in which we had sufficient information on response variables, power analyses were carried out to determine

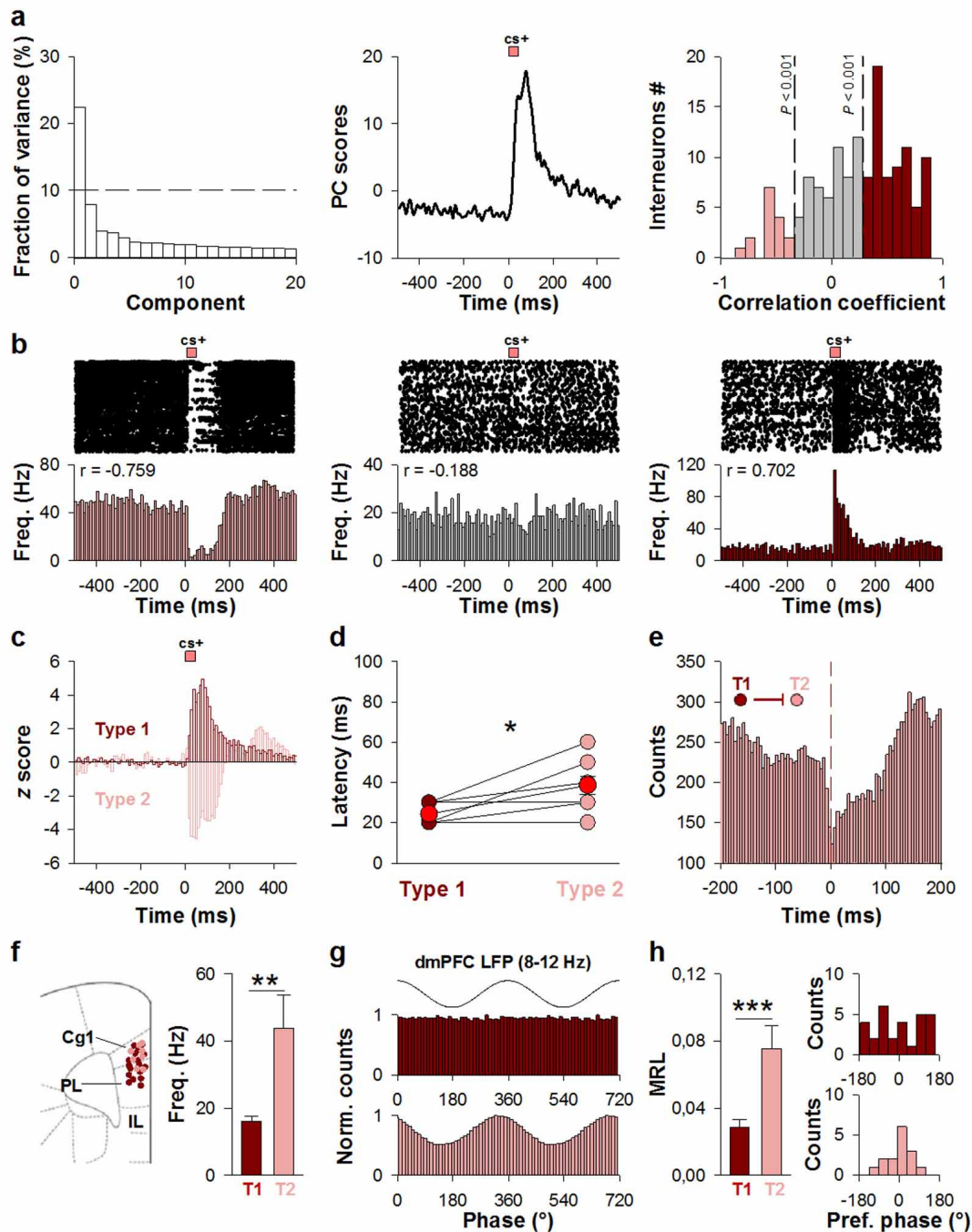
the number of mice needed. For studies in which the behavioural effect of the manipulation could not be pre-specified, such as optogenetic experiments, we used a sequential stopping rule (SSR). In essence this method enables null-hypothesis tests to be used in sequential stages, by analysing the data at several experimental points using *t*-tests. Usually the experiment started by testing only a few animals and if the *P* value was below 0.05, the investigator declared the effect significant and stopped testing. If the *P* value was greater than 0.36, the investigator stopped the experiment and retained the null hypothesis. For sample-size estimation using power analyses, we used an on-line power analysis calculator (G\*power3). For each analysis, sample size was determined using a power  $> 0.9$  and alpha error = 0.05. All tests were two sided. Power analyses were computed for matched pairs (differential conditioning protocol in which we used an internal control (Fig. 1) and pharmacological experiments (Extended Data Fig. 8)). In our behavioural experiments, a critical parameter is freezing level, and the numerical endpoint typically ranges between 50 and 70% freezing for CS<sup>+</sup> presentations immediately following auditory fear conditioning and between 10 and 30% freezing for CS<sup>−</sup> presentations. A minimum biologically significant difference in the mean values between CS<sup>−</sup> and CS<sup>+</sup> conditions (Fig. 1), or between CS<sup>+</sup> presentations before and after pharmacological treatment (Extended Data Fig. 8) is 1.5-fold. If we assume a standard deviation of 1.5 for a mean value of 60% freezing for CS<sup>+</sup> and 20% freezing for CS<sup>−</sup> or CS<sup>+</sup> after pharmacological treatment (which are realistic numbers), then a minimal  $n = 6$  (paired *t*-test) or  $n = 8$  (Wilcoxon signed-rank test) is needed to reject the null hypothesis with 90% probability. Sample size determination using SSRs analyses were used for optogenetic experiments in which it was not possible to determine a priori the effect of the optical manipulation. We used *P* values of 0.05 and 0.36 for lower and upper criterion. Using this strategy we ended up with an  $n$  comprising between 6 and 13 animals per group.



**Extended Data Figure 1 | Separation of putative principal neurons and putative interneurons.** **a**, Left panel, superimposed waveforms recorded from two different units. Right panel, spikes originating from individual units were sorted using three-dimensional principal-component analysis. **b**, Corresponding auto-correlograms, colour-coded as in **a**, displaying clear refractory periods. **c**, Among the population of dmPFC neurons displaying significant excitatory or inhibitory CS<sup>+</sup>-evoked responses ( $n = 493$ ), 71.2% were classified as putative principal neurons (PNs, blue circles,  $n = 351$ ) and 28.8% as putative interneurons (INs, red circles,  $n = 142$ ) using an unbiased unsupervised cluster-separation algorithm based on three electrophysiological properties: firing frequency, spike half-width and spike area under waveform (AUP) peak. Inset, average waveform of a representative PN and IN illustrating

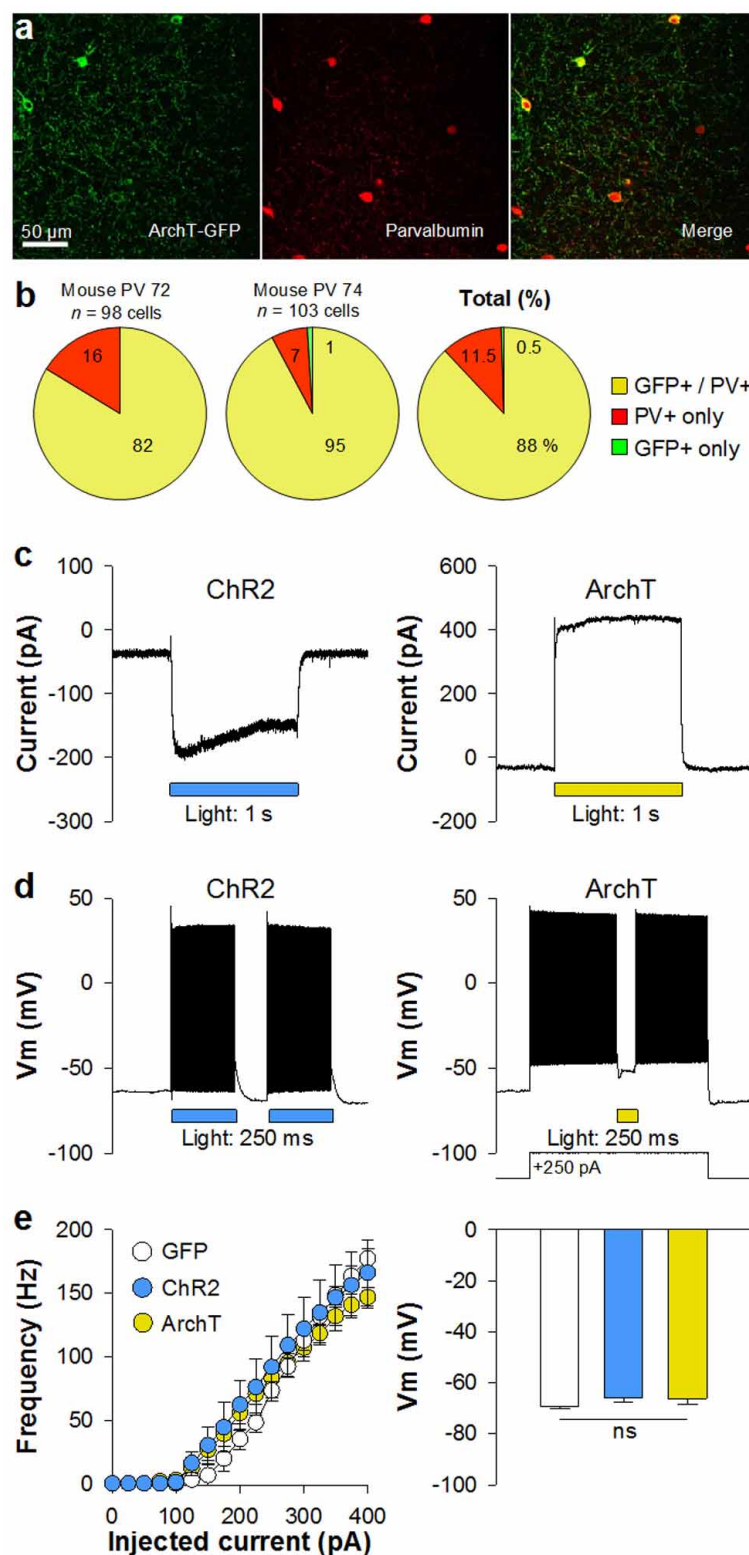
the methodology used to quantify spike width (SW) and the spike segment used to calculate the AUP. **d**, Top panel, representative cross-correlogram performed between a putative inhibitory IN and a non-identified neuron showing a short-latency, presumably monosynaptic, inhibitory interaction (7 pairs identified among putative INs, no inhibitory interaction among putative PNs). Bottom panel, representative cross-correlogram between a putative PN and a non-identified neuron showing a short-latency, possibly monosynaptic, excitatory interaction (20 pairs identified among PNs, no excitatory interaction from putative INs). Reference events correspond to the spikes of the pre-synaptic neuron (dashed line at time 0, bins of 0.5 ms). Grey circles represent neurons that were not tone-responsive.





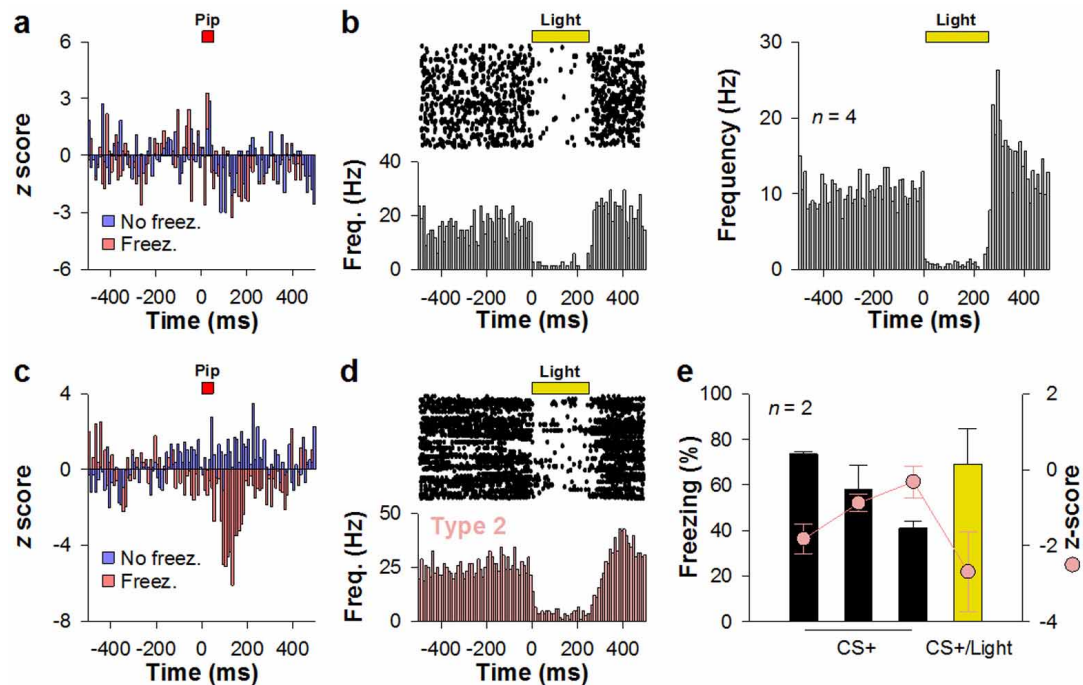
**Extended Data Figure 2 | CS<sup>+</sup>-evoked firing patterns and inhibitory interactions of putative INs.** **a**, Left panel, distribution of the fraction of variance for the 20 first principal components (PCs) obtained with principal component analysis (PCA). PC1, which accounted for more than 20% of variance of the data set, was used for the analysis. Middle panel, first principal-component coefficients representing the main firing pattern evoked by CS<sup>+</sup> (CS<sup>+</sup> onset at time 0) of the IN data set. Right panel, distribution of dmPFC IN correlation coefficients with PC1. The dashed lines indicate the levels of significance ( $P < 0.001$ ). Among the 142 INs, 83 (58.5%) displayed a significant positive ( $n = 68$ , 48%, dark red bars) or negative ( $n = 15$ , 10.6%, light red bars) correlation with PC1, whereas 41.5% INs ( $n = 59$ , grey bars) did not. **b**, Raster plots and PSTH of individual INs negatively correlated (left part, type 2 IN), not correlated (middle part) or positively correlated (right part, type 1 IN) with PC1. Type 1 INs were excited, whereas type 2 INs were inhibited by CS<sup>+</sup>. Bins of 10 ms. **c**, PSTH of all type 1 ( $n = 68$ ) and type 2 ( $n = 15$ ) INs illustrating the CS<sup>+</sup>-evoked responses (Post-FC, Ext. or Ret. Sessions, CS<sup>+</sup> 1–4). Bins of 10 ms. **d**, Individual (type 1 INs, dark red dots; type 2 INs, light red dots) and averaged (red dots) latencies of the first significant time bin ( $z$  score  $< -1.65$  or

$> +1.65$ ) following CS<sup>+</sup> for type 1 and type 2 INs recorded simultaneously ( $n = 7$  pairs recorded in 5 mice). CS<sup>+</sup>-evoked excitation in type 1 INs preceded CS<sup>+</sup>-evoked inhibition in type 2 INs (mean latency: type 1,  $24.3 \pm 2$  ms; type 2,  $38.6 \pm 4.6$  ms; paired  $t$ -test,  $*P < 0.05$ ). Error bars, mean  $\pm$  s.e.m. **e**, Cross-correlation analysis performed between a type 1 and a type 2 IN recorded simultaneously outside CS. The cross-correlogram shows a short latency, potentially monosynaptic, inhibitory interaction. Reference event, spikes of the type 1 IN (dashed line at time 0). Bins of 5 ms. **f**, Locations of recording sites and mean firing frequencies of type 1 (T1,  $n = 68$ ) and type 2 (T2,  $n = 15$ ) INs (Mann–Whitney test,  $**P < 0.01$ ; Cg1, anterior cingulate cortex; PL, prelimbic area; IL, infralimbic area). **g**, Firing modulation of representative type 1 and type 2 INs with dmPFC theta oscillations filtered in the 8–12-Hz range (12-min recordings). Bins of 10°. **h**, Mean strength of firing synchronization to local theta oscillations as measured with the mean resultant length (MRL) vector (left panel, Mann–Whitney test, type 1 versus type 2,  $***P < 0.001$ ) and distribution of the preferred phases (right panels) for type 1 and type 2 INs significantly phase-locked to theta oscillations (type 1,  $n = 29/68$ ; type 2,  $n = 15/15$ ).



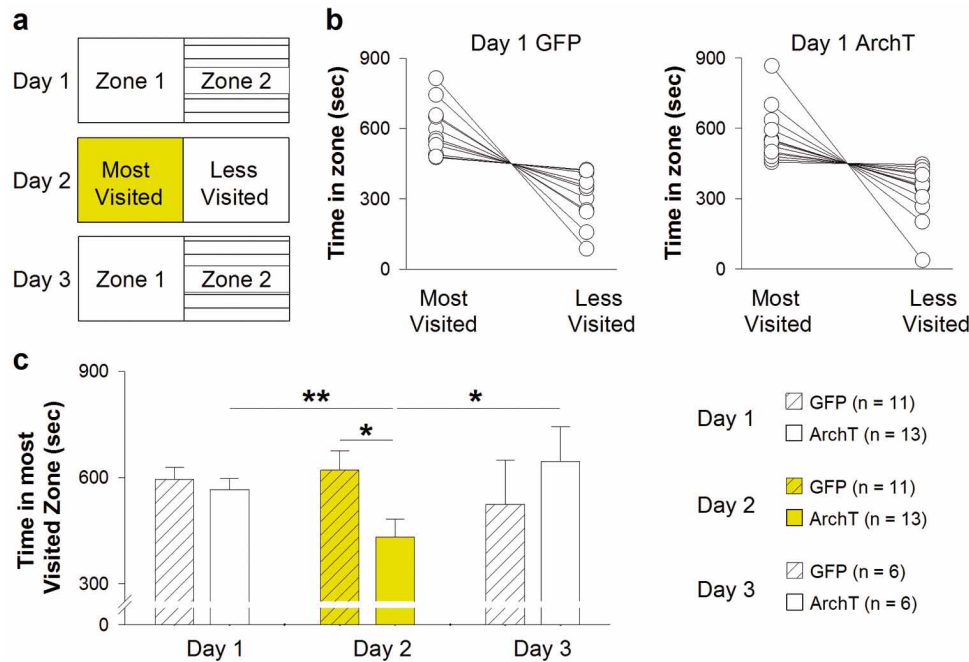
**Extended Data Figure 3 | Anatomical characterization of AAV-mediated ArchT-GFP expression in PV-IRES-Cre transgenic mice and electrophysiological characteristics of ArchT, ChR2 and GFP PV-IRES-Cre-infected PV neurons.** **a**, Representative confocal micrographs used for PV and GFP co-localization assessment. Left panel, ArchT-GFP labelled with anti-GFP Alexa 488; middle panel, PV immunofluorescence; right panel, merge. Single optical slices, in the same focal plane. **b**, Quantitative analysis of viral infection specificity and efficacy. Pie charts show the numbers of neurons positive for GFP and/or PV in two mice (left and middle charts) and averaged proportions (right chart). **c**, Representative ChR2- (left) and ArchT-evoked (right) currents recorded from PVINs with somata located in layer 2/3 of the

dmPFC (1 s light stimulation). **d**, Representative optically evoked action potential firing and inhibition of PVINs expressing ChR2 (left, 500-ms blue light pulses) or ArchT, respectively (right, 250-ms yellow light pulse during a 250-pA current pulse injection). **e**, Left panel, changes in firing frequency of PVINs expressing GFP (white dots,  $n = 7$ ), ChR2 (blue dots,  $n = 5$ ) or ArchT (yellow dots,  $n = 8$ ) upon injection of increasing current pulses (current pulses range, 0–400 pA). No significant differences were observed between groups. Right panel, resting membrane potentials of INs expressing GFP (white bar,  $n = 7$ ), ChR2 (blue bar,  $n = 5$ ) or ArchT (yellow bar,  $n = 8$ ). No significant differences were observed between groups (unpaired  $t$ -tests).



**Extended Data Figure 4 | Type 2 PVINs mediate conditioned fear responses.** **a**, z-score transformation of CS<sup>+</sup>-evoked firing of a non-type-2 IN for sound pips outside (No freez.) or inside (Freez.) freezing periods during the extinction session (CS<sup>+</sup> 1–12; No freez., 141 pips; Freez., 156 pips). This neuron was not classified as a type 1 or type 2 IN. **b**, Left panel, raster plot illustrating optogenetic identification of the same non-type-2 IN as ArchT-expressing (that is, PV-expressing). Right panel, mean z-score transformation of all non-type-2 INs identified as PV-expressing INs ( $n = 4$ ; light-pulse duration, 250 ms; 108 stimulation trials). **c**, z-score transformation of CS<sup>+</sup>-evoked firing of a type 2 IN for No freez. and Freez. periods during

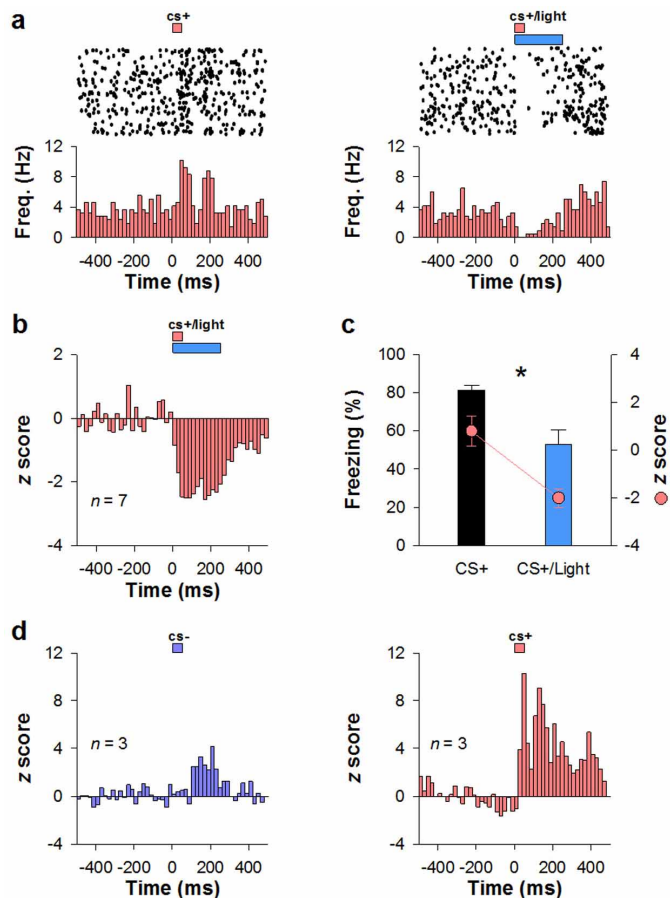
the extinction session (CS<sup>+</sup> 1–12; No freez., 141 pips; Freez., 156 pips). **d**, Raster plot illustrating optogenetic identification of the same type 2 IN as ArchT-expressing (that is, PV-expressing) (light-pulse duration, 250 ms; 108 stimulation trials). **e**, CS<sup>+</sup>-evoked changes in firing rate in two type 2 PVINs identified with optogenetic, and corresponding freezing scores of the two mice in which they were recorded (dots, mean z-score 150 ms post CS; bars, blocks of 4 CS<sup>+</sup> presentation each, both during the second extinction session; light-pulse duration, 250 ms; 108 stimulation trials). Light-induced inhibition of PV, including type 2 INs, reinstated freezing behaviour. Error bars indicate mean  $\pm$  s.e.m.



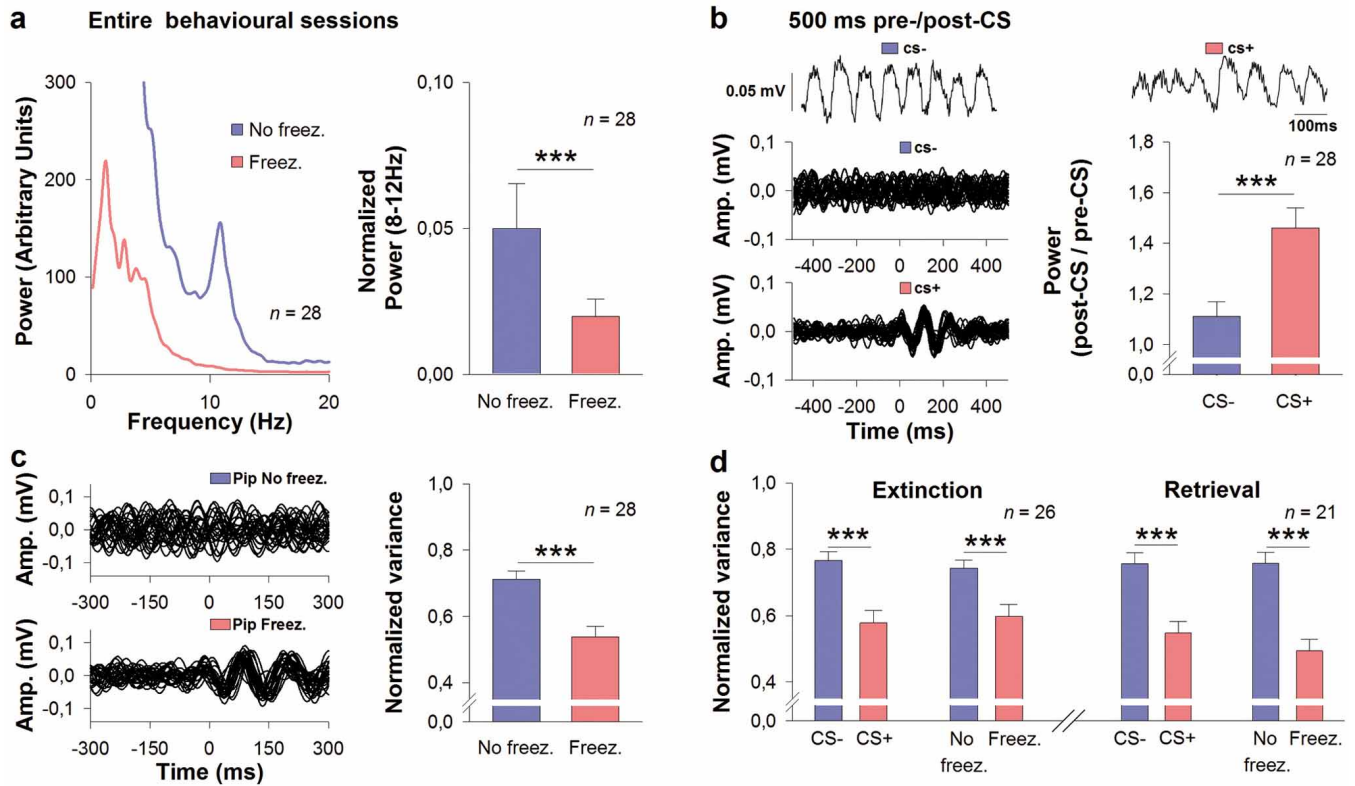
**Extended Data Figure 5 | Optogenetic inhibition of prefrontal PVINs induces place aversion.** **a**, On day 1, GFP- and ArchT-infected mice ( $n = 11$  and  $13$ , respectively) were exposed to a two-compartment place aversion apparatus during 15 min. Following pre-exposure, the most visited compartment was selected for each animal. On day 2, systematic yellow-light-induced inhibition of PVINs was triggered only in the most visited compartment during a 15-min exposure session. On day 3, GFP and ArchT infected mice ( $n = 6$  in both cases) were re-exposed to the place aversion apparatus during 15 min to evaluate the long-term effect of yellow-light stimulation during day 2. **b**, Time spent in the most and less visited compartments on day 1 for individual infected mice (GFP and ArchT). **c**, Average percentage of time spent in the most visited compartment on days 1,

2 and 3 for GFP- and ArchT-infected mice. A one-way analysis of variance (ANOVA) repeated measures performed on values from the GFP or the ArchT group revealed a significant effect only for the ArchT group (ArchT,  $F_{2,10} = 4.234$ ,  $P < 0.05$ ; GFP,  $F_{2,10} = 0.950$ ,  $P = 0.4191$ ). Post-hoc analysis revealed that on day 2, light inhibition of PVINs induced an aversion of the most visited compartment for ArchT infected animals in comparison to day 1 (ArchT mice, day 1 versus day 2, paired  $t$ -test,  $**P < 0.01$ ) and to GFP controls on day 2 (day 2, ArchT versus GFP, unpaired  $t$ -test,  $*P < 0.05$ ; 250-ms pulses delivered at 0.9 Hz). On day 3, ArchT mice did not avoid the most visited compartment any more (ArchT mice, day 2 versus day 3, unpaired  $t$ -test,  $*P < 0.05$ ). Error bars indicate mean  $\pm$  s.e.m.



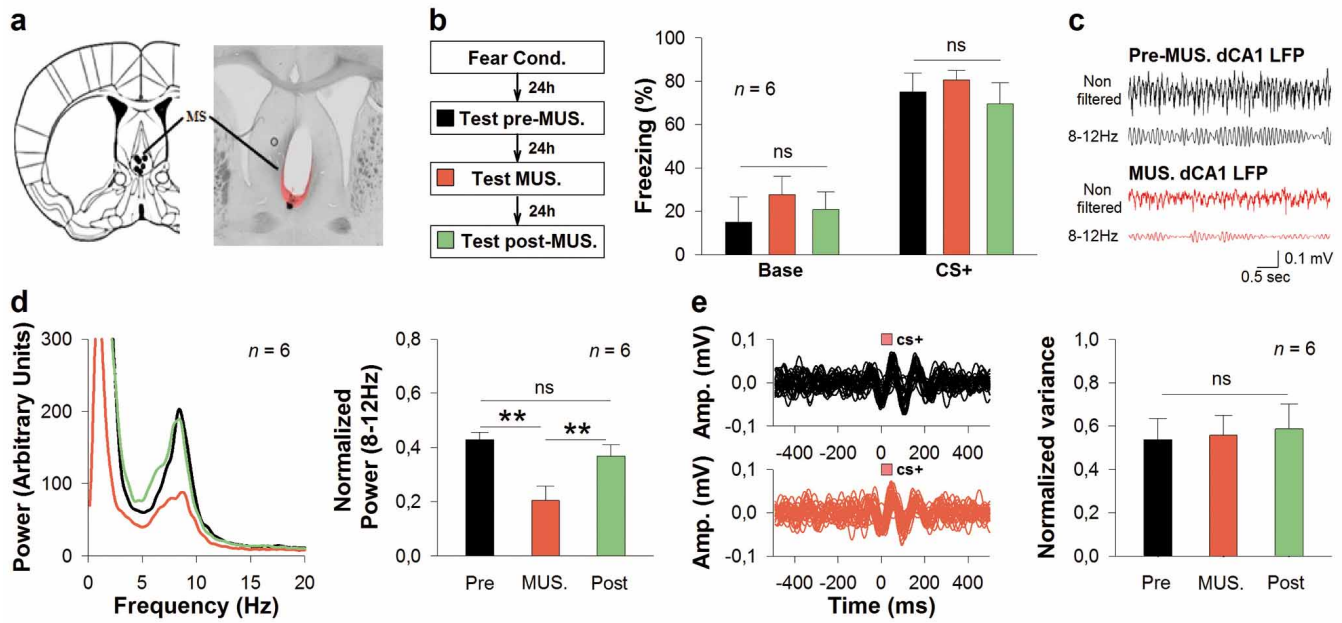


**Extended Data Figure 6 | Optogenetic activation of PVINs inhibits principal neurons and reduces freezing behaviour.** **a**, Raster plots and peristimulus time histograms illustrating the CS<sup>+</sup>-evoked excitation of a representative PN (left panel, Post FC, CS<sup>+</sup> presentations 1–4, 108 pips) and its blockade upon optogenetic-induced activation of PVINs (right panel, CS<sup>+</sup> presentations 5–8; light-pulse duration, 250 ms; 108 pips + stimulation trials) during the Post-FC session. **b**, z-score-transformed peristimulus time histogram showing PNs inhibition ( $n = 7$ ) following optogenetic-evoked activation of PVINs during CS<sup>+</sup> presentations (Post-FC session, CS<sup>+</sup> presentations 5–8; light-pulse duration, 250 ms; 108 stimulation trials). **c**, Freezing behaviour (bars,  $n = 3$  mice, block of 4 CS<sup>+</sup>) and CS<sup>+</sup>-evoked firing changes of PNs (red dots,  $n = 7$  neurons, mean z-score 100 ms post CS) before and in response to light-induced activation of PVINs during Post-FC sessions (light pulse duration, 250 ms; 108 stimulation trials; CS<sup>+</sup> 1–4 and 5–8, respectively). Optogenetic activation of PVINs inhibited PNs and reduced conditioned freezing behaviour (Wilcoxon signed-rank test,  $*P < 0.05$ ). **d**, z-score transformed peristimulus time histogram showing CS<sup>+</sup>-evoked excitation of PNs ( $n = 3$ ) exhibiting antidromic responses to BLA stimulations (Post-FC, CS<sup>-</sup> and CS<sup>+</sup> presentations 1–4, 108 pips each). These three neurons were included in the seven neurons for which CS<sup>+</sup>-evoked excitation was blocked by light excitation of PVINs (a and b). Error bars indicate mean  $\pm$  s.e.m.



**Extended Data Figure 7 | Transient amplitude increase and phase reset of local theta oscillations during fear expression.** **a**, Left panel, power spectrum of the non-filtered dmPFC LFPs recorded during Post-FC sessions ( $n = 28$  mice) for non-freezing (No freez.) and freezing (Freez.) periods showing a prominent 8–12-Hz component (that is, theta) only during non-freezing periods. Right panel, normalized theta power (8–12 Hz) for freezing and non-freezing periods during Post-FC sessions ( $n = 28$  mice, Wilcoxon signed-rank test  $***P < 0.001$ ). **b**, Top panels, non-filtered dmPFC LFP traces selected on the basis of prominent theta oscillations illustrating the transient increase in amplitude and phase reset of theta oscillations in response to CS<sup>+</sup> (Post-FC, 1 trial). Bottom-left panel, representative dmPFC 8–12-Hz LFP traces illustrating the phase reset and transient amplitude increase of theta oscillations in response to CS<sup>+</sup> or CS<sup>-</sup> presentations (Post-FC, 27 pips each). Bottom-right panel, average ratio of LFP theta power (500 ms post CS or 500 ms pre CS) in response to CS<sup>-</sup> and CS<sup>+</sup> pips. This analysis revealed a larger transient

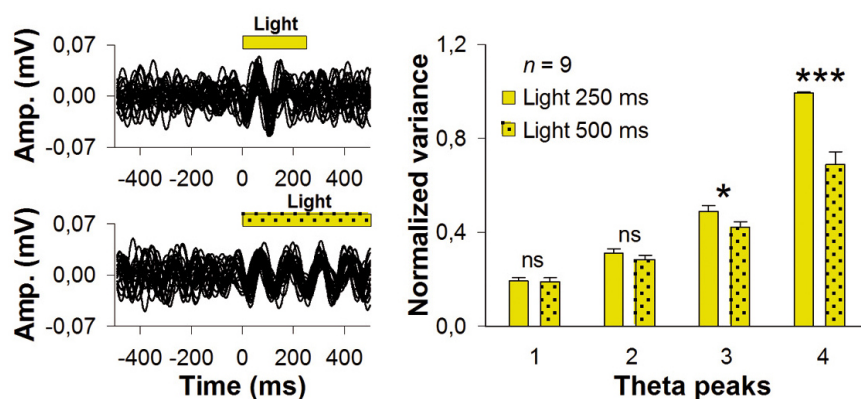
increase in LFP upon CS<sup>+</sup> presentations (Post-FC,  $n = 28$  mice, CS<sup>-</sup> versus CS<sup>+</sup>, paired  $t$ -test,  $***P < 0.001$ ). **c**, Left panel, representative dmPFC LFP traces filtered in the 8–12-Hz range, illustrating the phase resetting of theta oscillations during presentations of CS pips associated with no freezing or freezing behaviour (Post-FC, 27 pips). Right panel, quantification of the variance of the first theta peak occurrence following pip presentations in freezing and non-freezing periods (Post-FC,  $n = 28$  mice, No freez. versus Freez., paired  $t$ -test,  $***P < 0.001$ ). A small variance corresponds to a strong theta phase resetting. **d**, Quantification of the time variance of the first theta peak following CS<sup>-</sup> and CS<sup>+</sup> presentations or No freez. and Freez. periods for extinction and retrieval sessions (extinction, CS<sup>-</sup> presentations and CS<sup>+</sup> presentations 1–4,  $n = 28$  mice; retrieval, CS<sup>-</sup> and CS<sup>+</sup>,  $n = 21$  mice; CS<sup>-</sup> versus CS<sup>+</sup>, paired  $t$ -test,  $***P < 0.001$ ; No freez. versus Freez., paired  $t$ -test,  $***P < 0.001$ ). Error bars indicate mean  $\pm$  s.e.m.



**Extended Data Figure 8 | Targeted reversible inactivation of the medial septum does not block conditioned fear expression or theta phase resetting.**

**a**, Locations of injection sites in the medial septum (MS) and composite transmission light and epifluorescence micrograph showing the MS targeted injection of muscimol (MUS.) covalently bound to a fluorescent tag (right, dipyrromethene boron difluoride (BODIPY), red). **b**, Experimental design and mean freezing values of fear conditioned mice ( $n = 6$ ) before (Test Pre-MUS.), following (Test MUS.), and one day after (Post-MUS.) injections of MUS in the MS. Following fear conditioning, targeted inactivation of the MS had no effect on basal locomotor activity or CS<sup>+</sup>-evoked freezing responses (paired  $t$ -tests). **c**, Illustrative raw and filtered (8–120-Hz) LFP traces recorded in the dorsal CA1 (dCA1) before and following MUS injections in the MS. **d**, Left panel,

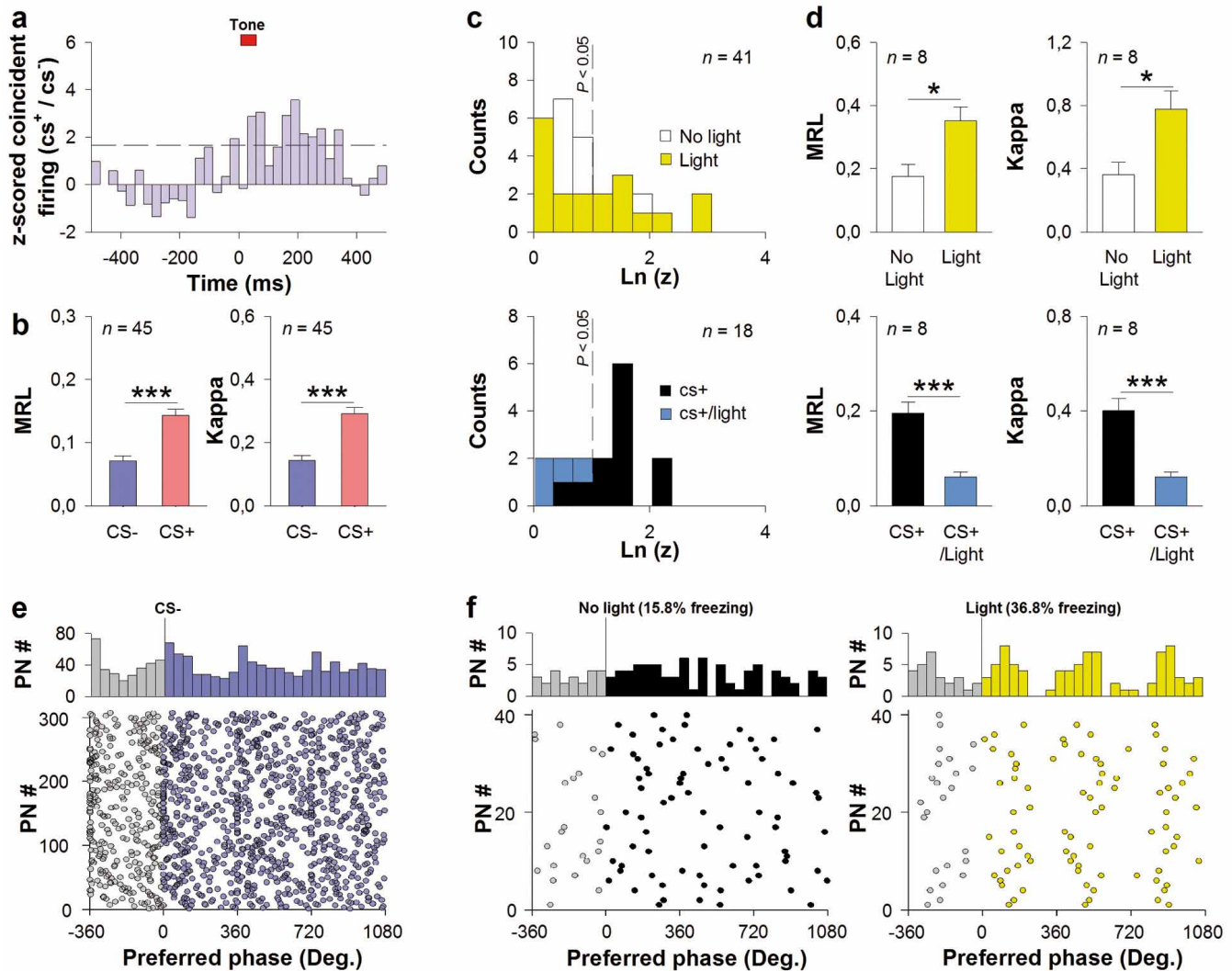
power spectra of dCA1 LFPs before, following and 1 day after MS inactivation. Right panel, quantification of dCA1 LFP power (8–12 Hz) before, during and after MS inactivation. MS inactivation significantly reduced dCA1 theta power ( $n = 6$  mice, Pre-MUS. versus MUS., paired  $t$ -test,  $**P < 0.01$ ; Post-MUS. versus MUS., paired  $t$ -test,  $**P < 0.01$ , ns, not significant). **e**, Resetting of prefrontal theta oscillations. Left panel, representative dmPFC LFP traces filtered in the 8–12-Hz range (Test-MUS., first CS<sup>+</sup>). Right panel, quantification of the time variance of the first theta peak following CS<sup>+</sup> presentations before, following and 1 day after MS inactivation (Pre-MUS., MUS., Post-MUS., CS<sup>+</sup> presentations 1–4, paired  $t$ -tests). MS inactivation had no effect on dmPFC theta phase resetting upon CS<sup>+</sup> presentations. Error bars indicate mean  $\pm$  s.e.m.



**Extended Data Figure 9 | Optogenetic inhibition of prefrontal PVINs resets local theta phase.** Left panel, representative 8–12-Hz filtered LFP traces, showing the resetting of local theta phase upon optogenetic inhibition of PVINs (top part, light-pulse duration, 250 ms, 27 stimulation trials; bottom part, light-pulse duration, 500 ms, 27 stimulation trials). Right panel, quantification

of the time variance of theta peaks (theta peaks 1–4) following presentations of 250-ms or 500-ms light pulses ( $n = 9$  mice, paired  $t$ -tests,  $*P < 0.05$ ,  $***P < 0.001$ , NS, not significant). Interestingly, dmPFC theta oscillations were precisely entrained for as long as PVINs were silenced. This suggests that inhibition from PVINs masks an oscillatory process in PNs.





**Extended Data Figure 10 | PVINs control principal-neuron theta phase locking and spiking synchronization.** **a**, Normalized averaged ratio of changes in coincident activity between pairs of PNs induced by CS<sup>+</sup> and CS<sup>-</sup> and corrected for changes in firing rate (Post-FC, Ext. or Ret. sessions;  $n = 975$  pairs from 308 PNs). Dashed line indicates significant z score ( $P < 0.05$ ). Bins of 30 ms. **b**, Mean vector length (MRL) and concentration of Von Mises fit ( $\kappa$ ) upon CS<sup>-</sup> or CS<sup>+</sup>, two measures of modulation strength in phase with theta oscillations (Post-FC, Ext. or Ret. Sessions). Only neurons significantly phase locked to theta and for which at least 50 spikes were recorded during CS<sup>+</sup> were included ( $n = 45$ ) (CS<sup>-</sup> versus CS<sup>+</sup>, Wilcoxon tests, \*\*\*  $P < 0.001$ ). Error bars indicate mean  $\pm$  s.e.m. CS<sup>+</sup> entrains a stronger locking of PN spikes to ongoing theta oscillations. Together with the precise timing between CS<sup>+</sup> onset (resetting) and subsequent theta cycles, this ensures robust, coincident and timed spiking of PNs. **c**, Distribution of log-transformed Rayleigh's test Z values of PN theta modulation before and upon light-induced inhibition (top,  $n = 41$  neurons) and light-induced activation (bottom,  $n = 18$  neurons) of PVINs (light-pulse duration, 250ms; 108 trials for each; yellow light, stimulation at the end of the behavioural session; blue light, stimulation during Post-FC session, CS<sup>+</sup> presentations 5–8). Dashed line indicates significant theta phase locking threshold ( $\ln(Z) = 1.1$ ,  $P = 0.05$ ). **d**, Theta modulation of PNs significantly phase locked to theta and displaying at least 15 spikes during No light and Light conditions. Modulation with local theta was measured with the MRL (top-left panel,  $n = 8$  neurons, yellow light stimulation, paired  $t$ -tests, No light versus Light, \*  $P < 0.05$ ; bottom-left panel,  $n = 8$  neurons, blue light stimulation, No

light versus Light, \*  $P < 0.05$ ) and  $\kappa$  (top-right panel,  $n = 8$  neurons, yellow light stimulation, paired  $t$ -tests, No light versus Light, \*\*\*  $P < 0.001$ ; bottom-left panel,  $n = 8$  neurons, blue light stimulation, No light versus Light, \*\*\*  $P < 0.001$ ). Error bars indicate mean  $\pm$  s.e.m. These results show that inhibiting PVINs is both sufficient to increase PNs' modulation with local theta, and necessary for theta entrainment of PNs evoked by CS<sup>+</sup>. **e**, Top panel, distribution of PNs' preferred theta phase ( $n = 308$ ) during cycles around CS<sup>-</sup>. The phases of LFPs were aligned to the first theta peak following CS<sup>-</sup> onset to mimic phase resetting of local theta (one theta cycle before, and three theta cycles following CS were included, bins of 45°). Bottom panel, distribution of individual PNs' preferred theta phases during theta cycles around CS<sup>-</sup> showing a synchronization of PNs around the peak of the LFP (Rayleigh's test for circular uniformity, first theta cycle post CS,  $P < 0.001$ , indicating that the circular distribution is not uniform). **f**, Top panel, distribution of PNs' preferred theta phase ( $n = 41$ ) during theta cycles outside light stimulation (left part, 15.8% freezing) and upon light-induced resetting of theta oscillations (right part, 36.8% freezing; one theta cycle before, and 3 theta cycles following CS were included, bins of 45°). Bottom panel, distributions of individual PNs' preferred theta phase outside and upon light stimulation. Despite a low number of neurons and a moderate freezing induced by light inhibition of dmPFC PVINs (36.8% freezing), this analysis revealed that light-induced reset of local theta oscillations promotes neuronal synchronization of PNs (Rayleigh's test for circular uniformity, first theta cycle post CS; Light,  $P < 0.001$ ; No light,  $P = \text{NS}$ ).

# Divergent angiocrine signals from vascular niche balance liver regeneration and fibrosis

Bi-Sen Ding<sup>1\*</sup>, Zhongwei Cao<sup>1\*</sup>, Raphael Lis<sup>1</sup>, Daniel J. Nolan<sup>1,2</sup>, Peipei Guo<sup>1</sup>, Michael Simons<sup>3</sup>, Mark E. Penfold<sup>4</sup>, Koji Shido<sup>1</sup>, Sina Y. Rabbany<sup>1,5</sup> & Shahin Rafii<sup>1</sup>

Chemical or traumatic damage to the liver is frequently associated with aberrant healing (fibrosis) that overrides liver regeneration<sup>1–5</sup>. The mechanism by which hepatic niche cells differentially modulate regeneration and fibrosis during liver repair remains to be defined<sup>6–8</sup>. Hepatic vascular niche predominantly represented by liver sinusoidal endothelial cells deploys paracrine trophogens, known as angiocrine factors, to stimulate regeneration<sup>9–15</sup>. Nevertheless, it is not known how pro-regenerative angiocrine signals from liver sinusoidal endothelial cells is subverted to promote fibrosis<sup>16,17</sup>. Here, by combining an inducible endothelial-cell-specific mouse gene deletion strategy and complementary models of acute and chronic liver injury, we show that divergent angiocrine signals from liver sinusoidal endothelial cells stimulate regeneration after immediate injury and provoke fibrosis after chronic insult. The pro-fibrotic transition of vascular niche results from differential expression of stromal-derived factor-1 receptors, CXCR7 and CXCR4 (refs 18–21), in liver sinusoidal endothelial cells. After acute injury, CXCR7 upregulation in liver sinusoidal endothelial cells acts with CXCR4 to induce transcription factor Id1, deploying pro-regenerative angiocrine factors and triggering regeneration. Inducible deletion of *Cxcr7* in sinusoidal endothelial cells (*Cxcr7*<sup>ΔAEC/ΔAEC</sup>) from the adult mouse liver impaired liver regeneration by diminishing Id1-mediated production of angiocrine factors<sup>9–11</sup>. By contrast, after chronic injury inflicted by iterative hepatotoxin (carbon tetrachloride) injection and bile duct ligation, constitutive FGFR1 signalling in liver sinusoidal endothelial cells counterbalanced CXCR7-dependent pro-regenerative response and augmented CXCR4 expression. This predominance of CXCR4 over CXCR7 expression shifted angiocrine response of liver sinusoidal endothelial cells, stimulating proliferation of desmin<sup>+</sup> hepatic stellate-like cells<sup>22,23</sup> and enforcing a pro-fibrotic vascular niche. Endothelial-cell-specific ablation of either *Fgfr1* (*Fgfr1*<sup>ΔAEC/ΔAEC</sup>) or *Cxcr4* (*Cxcr4*<sup>ΔAEC/ΔAEC</sup>) in mice restored the pro-regenerative pathway and prevented FGFR1-mediated maladaptive subversion of angiocrine factors. Similarly, selective CXCR7 activation in liver sinusoidal endothelial cells abrogated fibrogenesis. Thus, we demonstrate that in response to liver injury, differential recruitment of pro-regenerative CXCR7-Id1 versus pro-fibrotic FGFR1–CXCR4 angiocrine pathways in vascular niche balances regeneration and fibrosis. These results provide a therapeutic roadmap to achieve hepatic regeneration without provoking fibrosis<sup>1,2,4</sup>.

Despite the liver's capacity to undergo regeneration, chronic or overwhelming injury often causes liver fibrosis that culminates in cirrhosis and hepatic failure<sup>1–7</sup>. The integrated process of liver repair includes regeneration and wound healing characterized by synthesis of extracellular matrix proteins. Both processes are modulated by dynamic interplay between parenchymal hepatocytes and non-parenchymal cells<sup>7,22,24,25</sup>, including hepatic stellate cells<sup>1,23</sup>, inflammatory cells<sup>6,8</sup>, biliary epithelial cells and liver sinusoidal endothelial cells (LSECs)<sup>9,13,15,16</sup>.

As such, defining the multicellular crosstalk that balances regeneration and dysfunctional (maladaptive) healing<sup>5</sup> holds promise for designing treatment for liver diseases.

LSECs that line liver sinusoidal vasculature induce hepatic organogenesis in manners that extend beyond their passive role for metabolite delivery<sup>9,12,14,15</sup>. By deploying paracrine growth regulators, which we have defined as angiocrine factors, LSECs trigger regeneration of hepatocytes<sup>9–11,25,26</sup>. However, aberrant activation of LSECs in the context of chronic injury provokes fibrosis<sup>16,17</sup>. This dichotomy of LSEC niche function in mediating liver repair implies divergent angiocrine signals balance regeneration and fibrosis<sup>11</sup>. Therefore, we sought to decipher the mechanisms that subvert pro-regenerative capacity of LSECs to a pro-fibrotic state.

In response to tissue injury, cytokines and chemokines, such as stromal-derived factor (SDF)-1 (Cxcl12), are upregulated to initiate regeneration by switching on its receptors CXCR4 and CXCR7 (refs 18–21). Although CXCR4 activation both in haematopoietic and in vascular cells modulates angiogenesis and haematopoiesis, expression of another SDF-1 receptor, CXCR7, is mainly restricted to endothelial cells, with its function primarily believed to be pivotal in vascular patterning and tumour neo-angiogenesis. However, elucidating the mechanism by which the SDF-1 pathway orchestrates liver repair is hindered by the lack of cell-type-specific genetic models in defined settings of liver injuries.

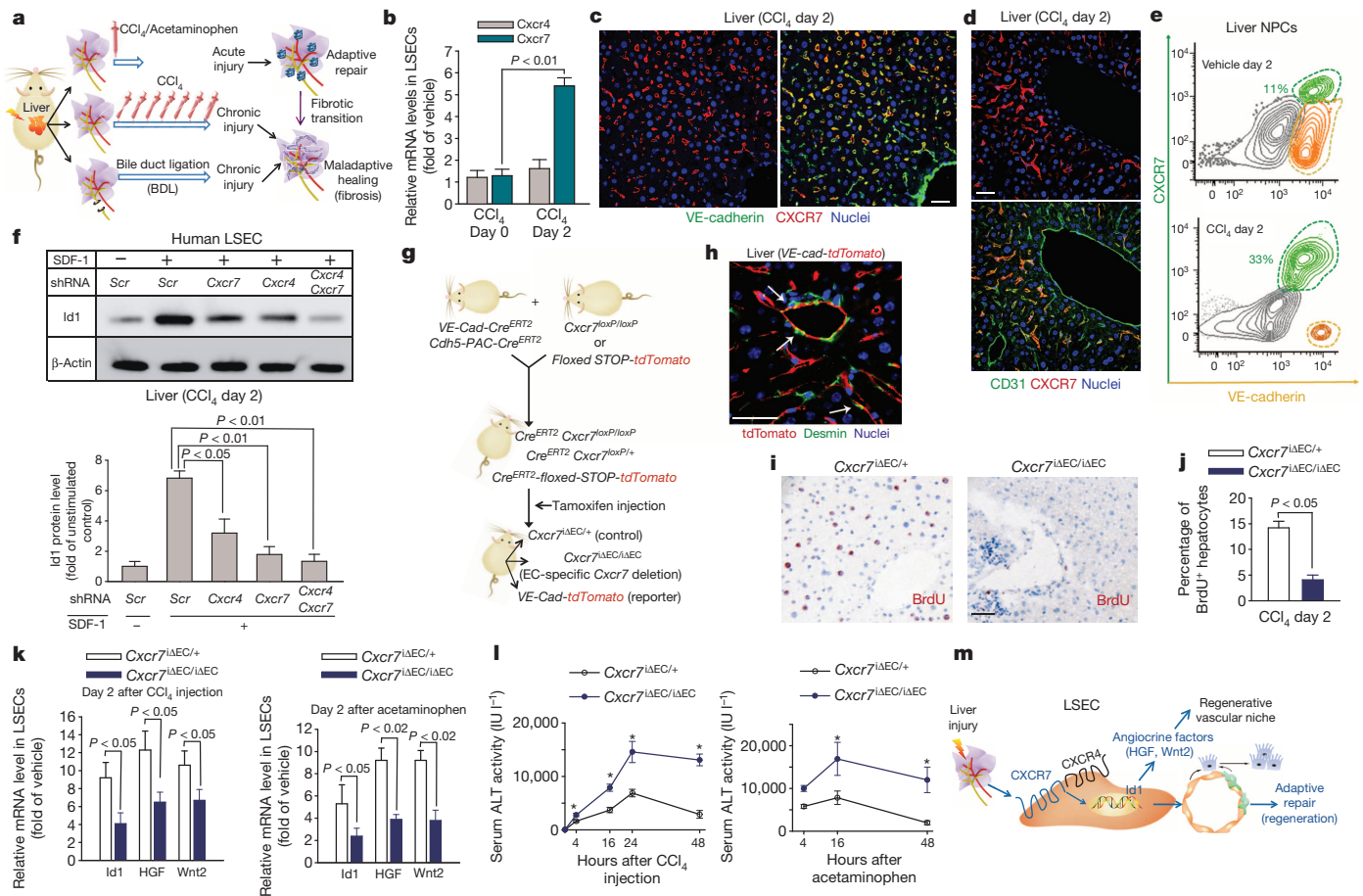
To unravel the divergent role of LSECs in modulating liver repair, we used a single injection of carbon tetrachloride (CCl<sub>4</sub>) and acetaminophen, which cause acute liver injury, as well as chronic injury models of repeated CCl<sub>4</sub> injection and bile duct ligation (BDL) (Fig. 1a). At day 2 after single CCl<sub>4</sub> injury, CXCR7 was upregulated specifically in vascular endothelial (VE)-cadherin<sup>+</sup> LSECs (Fig. 1b–e and Supplementary Fig. 1). By contrast, CXCR4 is broadly expressed by other cell types, and its expression remains relatively stable on LSECs after CCl<sub>4</sub> injury. Therefore, we have identified CXCR7 as an inducible LSEC-specific SDF-1 receptor in response to liver injury.

Our group<sup>9,10</sup> and others<sup>13</sup> have shown that after partial hepatectomy LSECs, by producing angiocrine factors such as Wnt2 and hepatocyte growth factor (HGF), elicit liver regeneration. Activation of transcription factor Id1 in LSECs was essential for this process<sup>9</sup>. SDF-1 induced Id1 upregulation in cultured human LSECs, which was abrogated by genetic silencing of either *Cxcr7* or *Cxcr4* (Fig. 1f and Supplementary Figs 2 and 3). CXCR7-selective agonist TC14012 similarly induced Id1 upregulation. Immunoprecipitation–western blot demonstrated that after SDF-1 stimulation, CXCR7 was associated with CXCR4 and β-arrestin in LSECs (Supplementary Fig. 4). Therefore, SDF-1 stimulates Id1 induction through enabling cooperation between CXCR7 and CXCR4 (refs 27 and 28).

To determine the contribution of CXCR7 in LSEC-mediated liver repair, we used a tamoxifen-inducible endothelial-cell-specific Cre<sup>ERT2</sup>

<sup>1</sup>Ansary Stem Cell Institute, Howard Hughes Medical Institute, Department of Genetic Medicine, Weill Cornell Medical College, New York, New York 10065, USA. <sup>2</sup>Angiocrine Bioscience, New York, New York 10065, USA. <sup>3</sup>Yale Cardiovascular Research Center, Yale University School of Medicine, New Haven, Connecticut 06510, USA. <sup>4</sup>ChemoCentryx, Inc., Mountain View, California 94043, USA. <sup>5</sup>Bioengineering Program, Hofstra University, Hempstead, New York 11549, USA.

\*These authors contributed equally to this work.



**Figure 1 | After acute liver injury, upregulation of SDF-1 receptor CXCR7 in LSECs stimulates angiocrine-mediated regeneration.** **a**, Liver injury models for studying the maladaptive transition of pro-regenerative LSEC function to a pro-fibrotic vascular niche. **b–e**, CXCR7 is specifically upregulated on VE-cadherin<sup>+</sup>CD31<sup>+</sup> LSECs after acute chemical injury. After injection of CCl<sub>4</sub>, CXCR7 and CXCR4 were determined in isolated LSECs (**b**), liver sections (**c**, **d**) and non-parenchymal cells (NPCs) (**e**). CXCR7 was expressed on LSECs but not large vessels; *n* = 5. Scale bar, 50 μm in Fig. 1; all data hereafter are presented as mean ± s.e.m. **f**, SDF-1 stimulation of human LSECs upregulates inhibitor of DNA binding 1 (Id1) protein, a transcription factor inducing production of pro-regenerative angiocrine factors<sup>9</sup>. Id1 stimulation by SDF-1 in primary human Factor VIII<sup>+</sup> LSECs was abrogated by silencing of *Cxcr4* and *Cxcr7* in LSECs; *n* = 5. **g**, **h**, Endothelial cell (EC)-specific inducible deletion of *Cxcr7* (*Cxcr7*<sup>ΔEC/ΔEC</sup>) in mice. Mice harbouring *loxP* sites flanking *Cxcr7* were crossed with a mouse line with endothelial-cell-specific *VE-cadherin* promoter-driven *Cre*<sup>ERT2</sup> (*VE-Cad-Cre*<sup>ERT2</sup>/*Cdh5-PAC-Cre*<sup>ERT2</sup>).

system to knockdown *Cxcr7* in the endothelial cells of adult mice (Fig. 1g). Mice harbouring *loxP* site-flanking *Cxcr7* were crossed with *VE-Cad-Cre*<sup>ERT2</sup>/*Cdh5-PAC-Cre*<sup>ERT2</sup> mice whereby endothelial-cell-specific *VE-cadherin* promoter drives *Cre*<sup>ERT2</sup>. Mice carrying tdTomato fluorescent protein following the floxed stop codon were used to exclude off-target effects of *VE-Cad-Cre*<sup>ERT2</sup> on other liver cell types. Tamoxifen injection specifically activated *Cre*<sup>ERT2</sup> activity in endothelial cells but not desmin-expressing stellate-like cells (Fig. 1h and Supplementary Fig. 5), demonstrating induced endothelial-cell-specific deletion of *Cxcr7* (*Cxcr7*<sup>ΔEC/ΔEC</sup>) in *VE-cad-Cre*<sup>ERT2</sup>/*Cxcr7*<sup>loxP/loxP</sup> mice. *Cxcr7*-haplodeficient adult mice (*Cxcr7*<sup>ΔEC/+</sup>) were used as control for Cre toxicity.

Compared with control mice, hepatocyte proliferation in *Cxcr7*<sup>ΔEC/ΔEC</sup> mice was significantly decreased after CCl<sub>4</sub> injury (Fig. 1i, j). Id1-dependent deployment of angiocrine factors HGF and Wnt2 from LSECs of *Cxcr7*<sup>ΔEC/ΔEC</sup> mice was reduced both after CCl<sub>4</sub>- and after acetaminophen-induced liver injuries (Fig. 1k). The extent of liver

Specificity of *VE-Cad-Cre*<sup>ERT2</sup>/*Cdh5-PAC-Cre*<sup>ERT2</sup> was validated in reporter mice carrying tdTomato protein following floxed stop codon. *Cxcr7* deletion or tdTomato expression in endothelial cells was induced by tamoxifen injection<sup>7</sup>. *Cxcr7*<sup>ΔEC/+</sup> mice served as control. Note the specific expression of tdTomato in endothelial cells but not desmin<sup>+</sup> stellate-like cells (**h**, white arrows). **i–l**, Impaired liver regeneration and enhanced liver injury in *Cxcr7*<sup>ΔEC/ΔEC</sup> mice after acute liver injury. Cell proliferation was determined by staining for bromodeoxyuridine (BrdU) incorporation (**i**, **j**). Expression of Id1 and pro-regenerative angiocrine factors, HGF and Wnt2, in LSECs was measured after CCl<sub>4</sub> and acetaminophen administration (**k**) and serum alanine aminotransferase (ALT) concentration was assessed to determine the degree of liver injury (**l**); *n* = 4. **m**, CXCR7 activation in LSECs triggers Id1-mediated production of pro-regenerative angiocrine factors. After acute liver injury, CXCR7 cooperates with CXCR4, inducing pro-regenerative Id1 pathway in LSECs and triggering angiocrine-mediated liver regeneration. Perivascular green cells denote stellate-like cells (refs 1–3, 22, 23).

injury, as determined by the concentration of serum alanine aminotransferase, was exacerbated (Fig. 1l and Supplementary Fig. 6). Thus, after liver injury, SDF-1 through activation of CXCR7<sup>+</sup> LSECs triggers an angiocrine response to initiate liver regeneration (Fig. 1m).

Although hepatocytes regenerate after acute liver injury, chronic liver damage more frequently leads to activation of myofibroblasts and causes fibrosis<sup>2,3,5</sup>. To address how the pro-regenerative angiocrine signals of LSECs are diverted to provoke this maladaptive healing, we used a mouse model of chronic liver injury by repeated CCl<sub>4</sub> injection<sup>29</sup> (Fig. 2a, b). The CXCR7-Id1 pathway in LSECs was counterbalanced by CXCR4 upregulation after chronic injury (Fig. 2c–e and Supplementary Figs 7 and 8). After repeated CCl<sub>4</sub> injection, protein concentrations of α-smooth muscle actin (SMA) and extracellular matrix protein collagen were augmented in *Cxcr7*<sup>ΔEC/ΔEC</sup> mice, compared with control mice (Fig. 2f–h and Supplementary Figs 9 and 10). Injection of CXCR7-specific agonist TC14012 reduced the upregulation of SMA and collagen I in control but not *Cxcr7*<sup>ΔEC/ΔEC</sup> mice (Fig. 2g, h and



Supplementary Fig. 11). Therefore, chronic liver injury interferes with pro-regenerative CXCR7-Id1 angiocrine pathway in LSECs and promotes fibrosis.

The requirement of CXCR7 in LSECs in resolving liver fibrosis was tested<sup>6,23</sup>. After three CCl<sub>4</sub> injections, SMA and collagen protein concentrations were enhanced in control mice; they peaked at day 8 after last injection and approached basal (vehicle-injected group) amount at day 20 (Fig. 2i–j and Supplementary Figs 12 and 13). By contrast, time-dependent resolution of liver injury was impaired in *Cxcr7*<sup>ΔEC/ΔEC</sup> mice. Id1 pathway inhibition by repeated injection of CCl<sub>4</sub> was prevented by CXCR7 agonist TC14012 (Fig. 2k). Therefore, in response to injury, the CXCR7 pathway in LSECs plays an indispensable role in stimulating regeneration and resolving fibrosis. After iterative stimuli, predominance of the CXCR4 pathway over the CXCR7-Id1 pathway in LSECs leads to fibrosis (Fig. 2l).

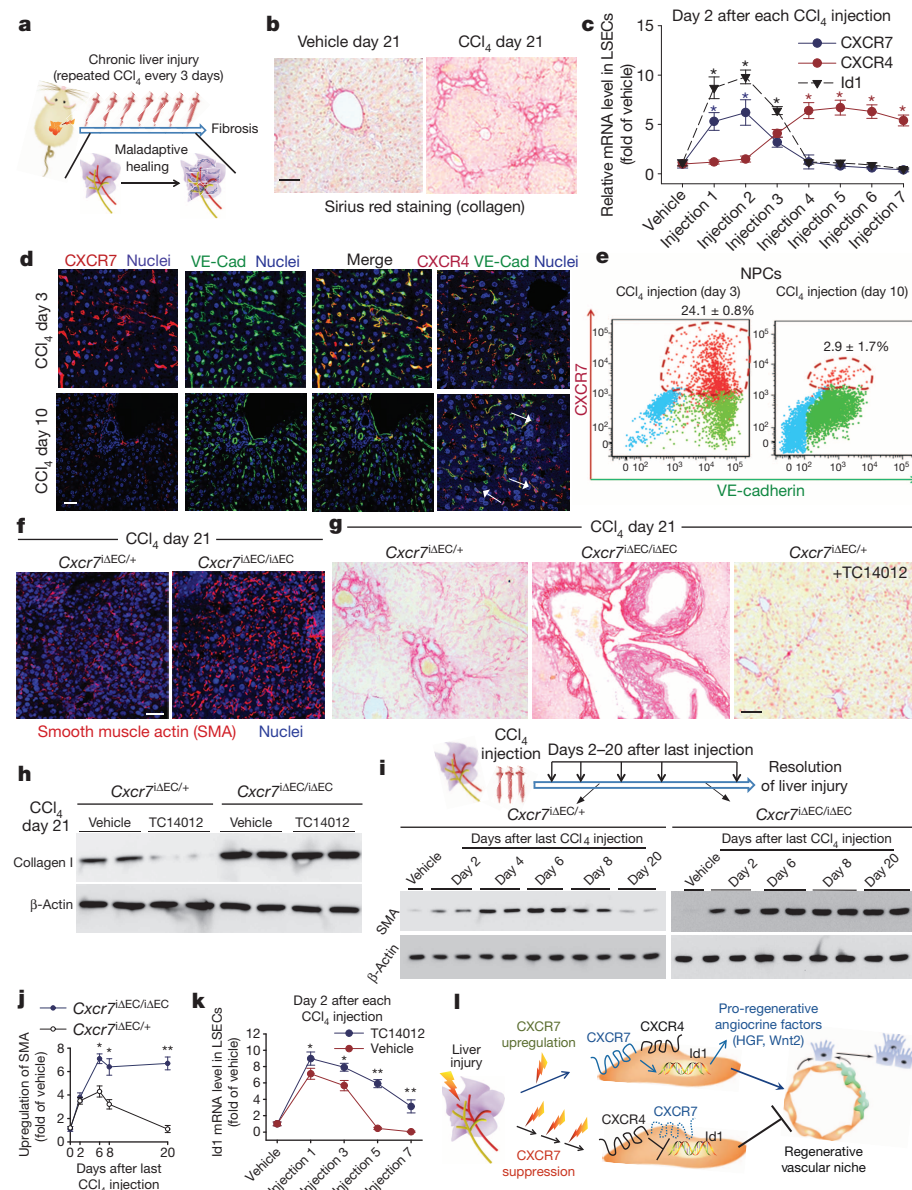
We then used BDL, a clinically relevant liver cholestasis model, to define how CXCR4 and CXCR7 modulate pro-fibrotic transition of LSECs (Fig. 3a). BDL induces biliary epithelial injury and causes continuous cholestasis and cirrhosis. After BDL, LSECs were invested by perisinusoidal desmin<sup>+</sup> stellate-like cells (Fig. 3b). Similar to repeated CCl<sub>4</sub> injuries, there was temporal upregulation of CXCR4 and

suppression of CXCR7-Id1 pathways in LSECs (Fig. 3c). SMA deposition in the liver of *Cxcr7*<sup>ΔEC/ΔEC</sup> mice was higher than that of control *Cxcr7*<sup>ΔEC/+</sup> mice (Fig. 3d–f). As such, loss of CXCR7 signalling in LSECs leads to fibrosis during BDL-induced cholestatic injury.

Human LSECs were then stimulated with angiogenic factors VEGF-A and FGF-2 to investigate the mechanism whereby CXCR4 expression is enhanced to dominate over the CXCR7 pathway. FGF-2, but not VEGF-A, increased CXCR4 messenger RNA (mRNA) and protein concentrations, and attenuated CXCR7 expression (Fig. 3g and Supplementary Fig. 14). Specific inhibition of mitogen-activated protein kinase (MAPK) blocked FGF-2-driven CXCR4 induction and CXCR7 inhibition in LSECs<sup>11</sup> (Fig. 3h). As a result, treatment of human LSECs by FGF-2 suppressed Id1 induction by SDF-1 (Fig. 3i, j and Supplementary Fig. 15), suggesting that FGF-2-induced CXCR4 upregulation and CXCR7 suppression in LSECs negate the Id1 pro-regenerative pathway.

To test how FGF-2 signalling modulates angiocrine response of LSECs, we examined the activation of the FGF-2 receptor FGFR1 on LSECs after BDL. There was a time-dependent upregulation and activation of FGFR1 (refs 16, 29, 30) concomitant with phosphorylation of MAPK (Erk1/2) in the injured VE-cadherin<sup>+</sup> LSECs (Fig. 3k, l and Supplementary Figs 16 and 17). Hence, cholestatic injury causes

**Figure 2 | Iterative hepatotoxic injury perturbs CXCR7 pro-regenerative pathway in LSECs and forces the generation of a pro-fibrotic vascular niche.** **a, b**, Mouse liver fibrosis is induced by repeated injection of CCl<sub>4</sub><sup>29</sup>. Sirius red staining was used to detect collagen in the injured liver. Scale bar, 50 μm in Fig. 2. **c–e**, Chronic liver injury suppresses CXCR7 pathway and upregulates CXCR4 expression in LSECs. Quantitative PCR (**c**), immunostaining (**d**) and flow cytometry (**e**) showed the abrogation of CXCR7-Id1 pathway in VE-cadherin (VE-Cad)<sup>+</sup> LSECs after chronic CCl<sub>4</sub> injury. CXCR4 is expressed in both endothelial cells and non-endothelial cells (white arrow). \**P* < 0.05 versus vehicle-treated mice; *n* = 5. **f–h**, CXCR7 activation in LSECs negates liver fibrosis. The extent of fibrosis was augmented in *Cxcr7*<sup>ΔEC/ΔEC</sup> mice, as evidenced by elevated hepatic amounts of α-SMA and collagen I (**g**, sirius red staining). CXCR7-selective agonist TC14012 reduced fibrosis in control but not *Cxcr7*<sup>ΔEC/ΔEC</sup> mice. **i, j**, Impaired resolution of liver injury in *Cxcr7*<sup>ΔEC/ΔEC</sup> mice. The amount of SMA in the injured liver was tested to assess the resolution of injury (**i**). Compared with control mice, the amount of SMA in *Cxcr7*<sup>ΔEC/ΔEC</sup> mice was enhanced after last CCl<sub>4</sub> injection and remained stable afterwards. The amount of collagen I was similarly assessed (Supplementary Fig. 13); \**P* < 0.05, \*\**P* < 0.01, versus control mice; *n* = 4. **k**, CXCR7 activation restores Id1 induction in chronically injured LSECs. TC14012 prevented Id1 suppression in LSECs during repeated CCl<sub>4</sub> injury. \**P* < 0.05, \*\**P* < 0.01, compared with vehicle group; *n* = 5. **l**, Interference with pro-regenerative CXCR7-Id1 pathway in LSECs causes pro-fibrotic transition of vascular niche. After injury, upregulation of the CXCR7-Id1 pathway in LSECs induces generation of hepatic-active angiocrine factors and stimulates regenerative regeneration. Chronic injury perturbs CXCR7-Id1 signalling, counteracting regeneration and provoking fibrogenesis.

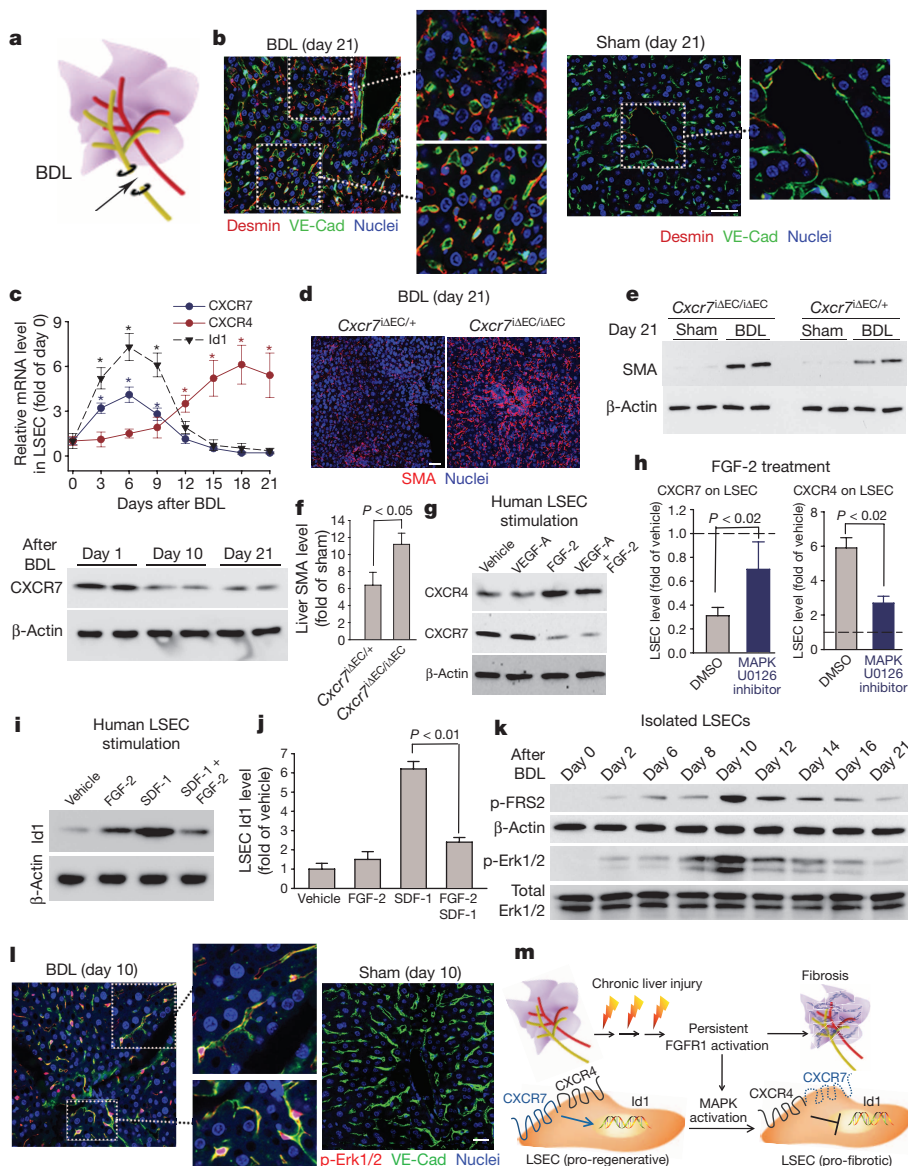




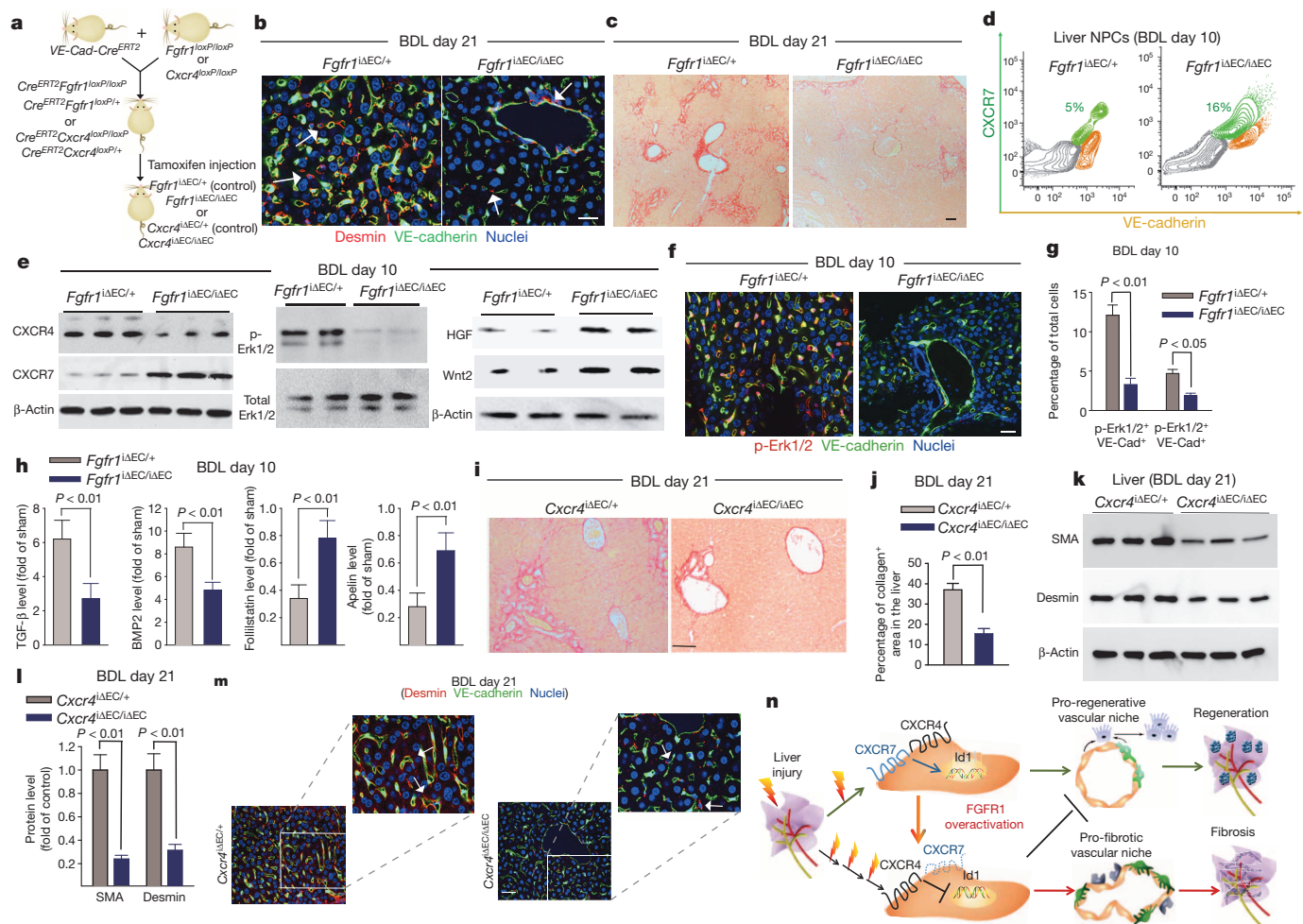
FGFR1-mediated MAPK activation in LSECs, resulting in CXCR4-dominated pro-fibrotic transition of angiocrine response during liver repair (Fig. 3m).

To elucidate the mechanism underlying the pro-fibrotic drift of LSEC niche, we conditionally ablated *Fgfr1* and *Cxcr4* in endothelial cells of adult mice (*Fgfr1*<sup>iAEC/iAEC</sup> and *Cxcr4*<sup>iAEC/iAEC</sup>) (Fig. 4a). After BDL, perisinusoidal expansion of desmin<sup>+</sup> cells, deposition of collagen and SMA, MAPK activation and CXCR7 suppression in LSECs of *Fgfr1*<sup>iAEC/iAEC</sup> mice were all prevented compared with control mice (Fig. 4b–g and Supplementary Fig. 18). CXCR7-dependent angiocrine expression of Wnt2 and HGF was restored in *Fgfr1*<sup>iAEC/iAEC</sup> mice (Fig. 4e). Therefore, endothelial-cell-specific deletion of *Fgfr1* in adult mice prevented the aberrant transition of LSECs into a pro-fibrotic state by BDL.

To unravel the altered angiocrine response in chronically injured LSECs, we isolated and analysed LSECs from BDL and sham-operated mice (Supplementary Fig. 19). In injured LSECs, there was significant upregulation of pro-fibrotic factors, including transforming growth factor (TGF)- $\beta$ , bone morphogenetic protein (BMP)2 and platelet-derived growth factor (PDGF)-C, concomitant with suppression of anti-fibrotic genes, such as follistatin and apelin. This divergent drift of angiocrine factor production in LSECs after BDL was diminished in *Fgfr1*<sup>iAEC/iAEC</sup> mice, as evidenced by restoration of anti-fibrotic genes and reduced expression of fibrotic factors (Fig. 4h).



**Figure 3 | Cholestatic liver injury by FGFR1 overactivation in LSECs shifts CXCR7-dependent pro-regenerative response to a CXCR4-dominated pro-fibrotic vascular niche.** **a, b**, Pro-fibrotic transition of LSECs caused by BDL-induced cholestatic injury. After BDL injury (**a**), most VE-cadherin<sup>+</sup> LSECs were covered by perisinusoidal desmin<sup>+</sup> fibroblasts (**b**, inset). By contrast, desmin<sup>+</sup> stellate-like cells were sparsely distributed in sham-operated liver. Scale bar, 50  $\mu$ m in Fig. 3. **c**, BDL suppresses CXCR7-Id1 pathway and upregulates CXCR4 expression in LSECs. Loss of CXCR7 protein in LSECs after BDL is shown at bottom panel; \* $P < 0.05$ , compared with the amount at day 0;  $n = 5$ . **d–f**, Liver fibrosis caused by BDL is exacerbated in *Cxcr7*<sup>iAEC/iAEC</sup> mice. After BDL, *Cxcr7*<sup>iAEC/iAEC</sup> mice showed higher hepatic concentrations of SMA protein than that of control mice;  $n = 4$ . **g–j**, FGF-2 through MAPK activation favours CXCR4 signalling in LSECs, counteracting CXCR7-Id1 pathway. FGF-2, but not VEGF-A, upregulated CXCR4, suppressed CXCR7 and inhibited SDF-1-dependent Id1 protein induction in LSECs. This FGF-2-mediated predominance of CXCR4 over CXCR7 was attenuated by MAPK inhibitor U0126; dashed line indicates the base level of protein expression in vehicle-treated LSECs, relative to FGF-2 stimulated cells. Note that DMSO served as control for MAPK inhibitor U0126.  $n = 5$ . **k, l**, Activation of FGFR1 and MAPK pathway in LSECs after BDL. There was time-dependent enhancement in phosphorylation/activation of FGFR1 downstream effector FRS2 (p-FRS2) and Erk1/2 (p-Erk1/2) in VE-cadherin<sup>+</sup> LSECs after BDL;  $n = 6$ . **m**, Constitutive FGFR1 signalling in LSECs by MAPK activation forces a CXCR4-dominated pro-fibrotic vascular niche. During chronic liver injury, FGFR1-mediated aberrant MAPK activation in LSECs upregulates CXCR4 and perturbs CXCR7-Id1 pathway. This predominance of FGFR1/CXCR4 activation in LSECs determines the pathological progression from adaptive (pro-regenerative) to maladaptive (pro-fibrotic) liver repair.



**Figure 4 | FGFR1 activation of CXCR4 in LSECs provokes pro-fibrotic angiocrine signals in liver repair.** **a**, endothelial-cell-specific inducible deletion of *Fgfr1* and *Cxcr4* (*Fgfr1<sup>ΔEC/ΔEC</sup>* and *Cxcr4<sup>ΔEC/ΔEC</sup>*) in adult mice. **b**, **c**, Reduced liver fibrosis in *Fgfr1<sup>ΔEC/ΔEC</sup>* mice. Compared with control mice, perisinusoidal enrichment of desmin<sup>+</sup> stellate-like cells (**b**, white arrow) and collagen deposition (**c**, sirius red staining) were diminished in the liver of *Fgfr1<sup>ΔEC/ΔEC</sup>* mice after BDL; *n* = 4; Scale bar, 50 μm in Fig. 4. **(d–g)** endothelial-cell-specific deletion of *Fgfr1* in mice prevents CXCR4-mediated maladaptive transition after BDL and restores regenerative angiocrine signals. In BDL-injured LSECs of *Fgfr1<sup>ΔEC/ΔEC</sup>* mice, CXCR7 suppression and CXCR4 upregulation (**d**) and Erk1/2 activation in VE-cadherin<sup>+</sup> LSECs (**e–g**) were reduced. This was accompanied by restored production of hepatic-active angiocrine factors HGF and Wnt2; *n* = 4. **h**, Pro-fibrotic production of angiocrine factors in LSECs is reduced in *Fgfr1<sup>ΔEC/ΔEC</sup>* mice. BDL instigated divergent production of angiocrine factors in LSECs, including upregulation of factors in BMP and TGF-β pathways and

suppression of anti-fibrotic genes such as follistatin and apelin (Supplementary Fig. 19). This pro-fibrotic drift of angiocrine factor in LSECs after BDL was mitigated in *Fgfr1<sup>ΔEC/ΔEC</sup>* mice; *n* = 4. **i–m**, Reduction of liver fibrosis in *Cxcr4<sup>ΔEC/ΔEC</sup>* mice after BDL. The extent of liver fibrosis after BDL was significantly lower in *Cxcr4<sup>ΔEC/ΔEC</sup>* mice than that of control mice, as evidenced by decreased deposition of collagen (**i**, **j**, sirius red staining), SMA protein (**k**, **l**), and perisinusoidal enrichment of desmin<sup>+</sup> stellate-like cells (**m**, white arrow); *n* = 5. **n**, Divergent angiocrine signals from LSECs balance liver regeneration and fibrosis. After acute liver injury, activation of CXCR7-Id1 pathway in LSECs stimulates production of hepatic-active angiocrine factors. By contrast, chronic injury causes persistent FGFR1 activation in LSECs that perturbs CXCR7-Id1 pathway and favours a CXCR4-driven pro-fibrotic angiocrine response, thereby provoking liver fibrosis. Therefore, in response to injury, differentially primed LSECs deploy divergent angiocrine signals to balance liver regeneration and fibrosis.

During chronic liver injury, loss of CXCR7 and upregulation of CXCR4 in LSECs causes progression to fibrosis. Indeed, the critical function of CXCR7 activation in promoting regeneration and counteracting fibrosis is evidenced by the attenuated fibrosis by selective activation of CXCR7 in LSECs, as well as by impaired regeneration (Fig. 1) and enhanced fibrogenesis (Figs 2 and 3) in *Cxcr7<sup>ΔEC/ΔEC</sup>* mice.

The shift in SDF-1 signalling from the CXCR7-dependent pro-regenerative response to a CXCR4-dominated pro-fibrotic function in LSECs is due to persistent FGFR1 that drives chronic MAPK activation<sup>11,16,29,30</sup> (Fig. 3). We used an inducible endothelial-cell-specific mouse genetic deletion system to demonstrate the contribution of FGFR1–CXCR4 pathway in the pro-fibrotic drift of LSECs. After chronic liver injury, the enhanced CXCR4 expression relative to CXCR7 in LSECs was prevented in *Fgfr1<sup>ΔEC/ΔEC</sup>* mice, and both *Fgfr1<sup>ΔEC/ΔEC</sup>* and *Cxcr4<sup>ΔEC/ΔEC</sup>* mice were resistant to fibrosis (Fig. 4). Therefore,

activation of CXCR7-Id1 in LSECs upon injury triggers and safeguards production of hepatogenic angiocrine factors, whereas persistent FGFR1–CXCR4 activation by chronic stimuli converts LSECs to a pro-fibrotic niche. This aberrant angiocrine response of LSECs by the FGFR1–CXCR4 pathway causes activation and expansion of desmin<sup>+</sup> stellate-like cells. Elucidating how hepatic stellate cells reciprocally modulate phenotypic and functional contributions of LSECs in liver repair remains to be investigated<sup>12,23</sup>. Moreover, whether activation of CXCR7 and CXCR4 on LSECs regulates inflammatory responses, such as recruitment of macrophages that could potentially modulate liver regeneration and fibrosis<sup>6,8</sup>, needs to be studied.

In summary, we have found that differentially activated LSECs supply divergent angiocrine signals for liver repair. Selective activation of CXCR7 in LSECs is instrumental in shepherding angiocrine-mediated regeneration. Perturbation of the CXCR7 pathway by constitutive FGFR1



activation diverts SDF-1 signalling in LSECs to a CXCR4-dominated maladaptive (pro-fibrotic) angiocrine response. Taken together, we demonstrate that endothelial cells are not just inert cellophane conduits delivering metabolites but also establish a vascular niche that instructively dictates regeneration and healing. Identifying molecular pathways orchestrating divergent angiocrine responses in the hepatic vascular niche will lay the foundation for therapeutic strategy that ensures liver repair without causing fibrosis.

## METHODS SUMMARY

Generation of endothelial-cell-specific inducible gene deletion was performed by treating *VE-Cad-Cre<sup>ERT2</sup>/Cdh5-PAC-Cre<sup>ERT2</sup>*-harbouring mice with tamoxifen<sup>10</sup>. Mice carrying *loxP* sites flanking *Cxcr7* were provided by ChemoCentryx (generated by L. Gan (University of Rochester)). Floxed *Cxcr4* mice were offered by Y.-R. Zou (the Feinstein Institute for Medical Research, Manhasset, New York). Sex-, age- and weight-matched littermate animals with indicated genotypes were used and compared in all experimental groups. All animal experiments were performed under the guidelines set by the Institutional Animal Care and Use Committee at Weill Cornell Medical College.

Single and repeated injections of CCl<sub>4</sub> were used to induce acute and chronic liver injuries, respectively<sup>29</sup>. To perform BDL, the common bile duct was ligated and severed by incision between ligation. Livers were collected for the analysis of fibrogenesis, including collagen deposition by Sirius red staining, and desmin protein by immunoblot (Abcam). TC14012 (R&D) was injected intraperitoneally into the mice after CCl<sub>4</sub> injury every other day at 30 mg kg<sup>-1</sup>. LSECs were isolated from mice as previously described<sup>9</sup>. Human LSECs were obtained from ScienCell<sup>TM</sup> Research Laboratories. Tissues were collected and cryopreserved as previously described<sup>9,10</sup>. Antibodies characterizing liver vasculature and fibroblasts were incubated with cryosections<sup>9</sup>. Flow cytometry analysis of LSECs was performed on liver non-parenchymal cells<sup>9</sup>. The sample size (*n*) of each experimental group is described in each corresponding figure legend, and all experiments were repeated at least three times. Statistical analysis used two-way analysis of variance. All data are presented as mean ± s.e.m.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 April; accepted 16 September 2013.

Published online 20 November 2013.

- Friedman, S. L., Sheppard, D., Duffield, J. S. & Violette, S. Therapy for fibrotic diseases: nearing the starting line. *Sci. Transl. Med.* **5**, 167sr1 (2013).
- Battaller, R. & Brenner, D. A. Liver fibrosis. *J. Clin. Invest.* **115**, 209–218 (2005).
- Iredale, J. P. Models of liver fibrosis: exploring the dynamic nature of inflammation and repair in a solid organ. *J. Clin. Invest.* **117**, 539–548 (2007).
- Gurtner, G. C., Werner, S., Barrandon, Y. & Longaker, M. T. Wound repair and regeneration. *Nature* **453**, 314–321 (2008).
- Wynn, T. A. Common and unique mechanisms regulate fibrosis in various fibroproliferative diseases. *J. Clin. Invest.* **117**, 524–529 (2007).
- Duffield, J. S. *et al.* Selective depletion of macrophages reveals distinct, opposing roles during liver injury and repair. *J. Clin. Invest.* **115**, 56–65 (2005).
- Diehl, A. M. Neighborhood watch orchestrates liver regeneration. *Nature Med.* **18**, 497–499 (2012).
- Boulter, L. *et al.* Macrophage-derived Wnt opposes Notch signaling to specify hepatic progenitor cell fate in chronic liver disease. *Nature Med.* **18**, 572–579 (2012).
- Ding, B. S. *et al.* Inductive angiocrine signals from sinusoidal endothelium are required for liver regeneration. *Nature* **468**, 310–315 (2010).
- Ding, B. S. *et al.* Endothelial-derived angiocrine signals induce and sustain regenerative lung alveolarization. *Cell* **147**, 539–553 (2011).
- Kobayashi, H. *et al.* Angiocrine factors from Akt-activated endothelial cells balance self-renewal and differentiation of haematopoietic stem cells. *Nature Cell Biol.* **12**, 1046–1056 (2010).
- Sakaguchi, T. F., Sadler, K. C., Crosnier, C. & Stainier, D. Y. Endothelial signals modulate hepatocyte apical-basal polarization in zebrafish. *Curr. Biol.* **18**, 1565–1571 (2008).

- Wang, L. *et al.* Liver sinusoidal endothelial cell progenitor cells promote liver regeneration in rats. *J. Clin. Invest.* **122**, 1567–1573 (2012).
- Matsumoto, K., Yoshitomi, H., Rossant, J. & Zaret, K. S. Liver organogenesis promoted by endothelial cells prior to vascular function. *Science* **294**, 559–563 (2001).
- LeCouter, J. *et al.* Angiogenesis-independent endothelial protection of liver: role of VEGFR-1. *Science* **299**, 890–893 (2003).
- Huebert, R. C. *et al.* Aquaporin-1 facilitates angiogenic invasion in the pathological neovasculature that accompanies cirrhosis. *Hepatology* **52**, 238–248 (2010).
- Zeisberg, E. M. *et al.* Endothelial-to-mesenchymal transition contributes to cardiac fibrosis. *Nature Med.* **13**, 952–961 (2007).
- Miao, Z. *et al.* CXCR7 (RDC1) promotes breast and lung tumor growth *in vivo* and is expressed on tumor-associated vasculature. *Proc. Natl Acad. Sci. USA* **104**, 15735–15740 (2007).
- Yu, S., Crawford, D., Tsuchihashi, T., Behrens, T. W. & Srivastava, D. The chemokine receptor CXCR7 functions to regulate cardiac valve remodeling. *Dev. Dyn.* **240**, 384–393 (2011).
- Sierro, F. *et al.* Disrupted cardiac development but normal hematopoiesis in mice deficient in the second CXCL12/SDF-1 receptor, CXCR7. *Proc. Natl Acad. Sci. USA* **104**, 14759–14764 (2007).
- Tachibana, K. *et al.* The chemokine receptor CXCR4 is essential for vascularization of the gastrointestinal tract. *Nature* **393**, 591–594 (1998).
- Armulik, A., Genove, G. & Betsholtz, C. Pericytes: developmental, physiological, and pathological perspectives, problems, and promises. *Dev. Cell* **21**, 193–215 (2011).
- Troeger, J. S. *et al.* Deactivation of hepatic stellate cells during liver fibrosis resolution in mice. *Gastroenterology* **143**, 1073–1083 (2012).
- Zaret, K. S. & Grompe, M. Generation and regeneration of cells of the liver and pancreas. *Science* **322**, 1490–1494 (2008).
- Woo, D. H. *et al.* Direct and indirect contribution of human embryonic stem cell-derived hepatocyte-like cells to liver repair in mice. *Gastroenterology* **142**, 602–611 (2012).
- Hoehme, S. *et al.* Prediction and validation of cell alignment along microvessels as order principle to restore tissue architecture in liver regeneration. *Proc. Natl Acad. Sci. USA* **107**, 10371–10376 (2010).
- Decailot, F. M. *et al.* CXCR7/CXCR4 heterodimer constitutively recruits  $\beta$ -arrestin to enhance cell migration. *J. Biol. Chem.* **286**, 32188–32197 (2011).
- Rajagopal, S. *et al.* Beta-arrestin- but not G protein-mediated signaling by the “decoy” receptor CXCR7. *Proc. Natl Acad. Sci. USA* **107**, 628–632 (2010).
- Yu, C. *et al.* Role of fibroblast growth factor type 1 and 2 in carbon tetrachloride-induced hepatic injury and fibrogenesis. *Am. J. Pathol.* **163**, 1653–1662 (2003).
- Bohm, F. *et al.* FGF receptors 1 and 2 control chemically induced injury and compound detoxification in regenerating livers of mice. *Gastroenterology* **139**, 1385–1396 (2010).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We are grateful to T. Hla for reading the manuscript, R. Berahovich and K. Eggen for their suggestions on CXCR7 signalling for liver repair, R. H. Adams for providing mouse inducible endothelial-cell-specific Cre (*Cdh5-PAC-Cre<sup>ERT2</sup>*) line and Y.-R. Zou for offering the floxed *Cxcr4* mouse line. B.-S.D. is supported by a National Scientist Development Grant from the American Heart Association (number 12SDG1213004) and a Druckenmiller Fellowship from the New York Stem Cell Foundation. S.R. is supported by the Ansary Stem Cell Institute, the Howard Hughes Medical Institute, Empire State Stem Cell Board and New York State Department of Health grants (C024180, C026438, C026878, C028117), National Heart, Lung, and Blood Institute grants R01HL097797, R01HL119872 and RC2HL101846, National Institute of Diabetes and Digestive and Kidney Diseases grant R01DK095039, National Cancer Institute grant U54CA163167, Qatar National Priorities Research Foundation grant NPRP08-663-3-140 and the Qatar Foundation BioMedical Research Program. R.L., P.G. and K.S. are supported by Empire State Stem Cell Board and New York State Department of Health training grants (C026878). D.N. is supported by the Tri-Institutional Weill Cornell Starr Stem Cell Scholar Program. M.S. is supported by National Heart, Lung, and Blood Institute grant R01 HL053793.

**Author Contributions** B.-S.D. and Z.C. conceived the project, performed experiments and wrote the paper. R.L., D.N. and P.G. performed the experiments and analysed the data. M.E.P., M.S., K.S. and S.Y.R. interpreted the data. S.R. designed the project, analysed the data and wrote the paper. All authors commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.R. ([srafii@med.cornell.edu](mailto:srafii@med.cornell.edu)) or B.-S.D. ([bid2004@med.cornell.edu](mailto:bid2004@med.cornell.edu)).

## METHODS

**Endothelial cell (EC)-specific gene deletion strategy.** Inducible endothelial-cell-specific gene deletion was achieved by treating *VE-Cadherin-Cre<sup>ERT2</sup>* harbouring mice with tamoxifen<sup>31</sup>. *Cxcr4<sup>loxP/loxP</sup>* mice were previously described<sup>32</sup>. The *Cre<sup>+</sup>* mice were treated with tamoxifen at a dose of 250 mg kg<sup>-1</sup> intraperitoneally for 6 days interrupted for 3 days after the third dose. After 3 weeks of tamoxifen treatment, deletion of target genes in LSECs was corroborated by quantitative PCR and immunoblot analysis. All animal experiments were performed under the guidelines set by the Institutional Animal Care and Use Committee at Weill Cornell Medical College, using sex-, age- and weight-matched littermate animals.

**Liver injury and fibrosis models.** Single and repeated injections of CCl<sub>4</sub> were used to induce acute and chronic liver injuries, respectively, as previously described<sup>29</sup>. CCl<sub>4</sub> was diluted in oil to yield a final concentration of 40% (0.64 mg ml<sup>-1</sup>) and injected into mice at 1.6 mg kg<sup>-1</sup> body mass. Eight- to ten-week-old mice were subjected to BDL. Acute liver injury was also induced in mice by intraperitoneal injection of 400 mg kg<sup>-1</sup> acetaminophen. To perform BDL, mice were subjected to a mid-abdominal incision 3 cm long, under general anaesthesia. The common bile duct was ligated in two adjacent positions approximately 1 cm from the porta hepatis. The duct was then severed by incision between the two sites of ligation.

To selectively activate CXCR7, the agonist TC14012 (R&D Systems) was intraperitoneally injected into the mice after CCl<sub>4</sub> injury or BDL injury every other day at 30 mg kg<sup>-1</sup>. At indicated time points, mice were killed and whole liver tissues were collected for the analysis of fibrogenesis, including collagen deposition by Sirius red staining and deposition of SMC and collagen I protein detected by immunoblot (Abcam).

**Isolation and culture of mouse liver cells.** Liver cells were isolated from mice by a two-step collagenase perfusion technique with modifications, as previously described<sup>9</sup>. Briefly, the liver was perfused with Liver Perfusion Medium (Invitrogen), and dissociated by Liver Digest Medium (Invitrogen). The non-parenchymal cells were fractionated with percoll gradient centrifugation with 75% stock Percoll solution and 35% stock Percoll solution. LSEC fraction was isolated by mouse LSEC binding magnetic beads (Miltenyi) and Dynabeads Magnetic Beads conjugated with anti mouse-VEGFR3 antibody (Imclone, NY)<sup>9</sup>. Expression of Id1, CXCR4, CXCR7 and FGFR1 protein and mRNA were determined from isolated LSECs<sup>11</sup>. For detection of FRS-2 phosphorylation in LSECs, mice were perfused with phosphatase inhibitor before collecting the tissues (Pierce)<sup>11</sup>.

**Culture and stimulation of human LSECs.** Human LSECs obtained from ScienCell<sup>TM</sup> Research Laboratories were cultured following vendor's instruction. The expression of CXCR7, VE-cadherin, vWF and factor VIII was validated by immunostaining or flow cytometric analysis. To selectively knockdown *Cxcr4*, *Cxcr7* in LSECs, shRNA Lentiviruses were generated by cotransfecting 15 µg of shuttle lentiviral vector containing *target gene* or scrambled shRNA, 3 µg of pENV/VSV-G, 5 µg of pRRE and 2.5 µg of pRSV-REV in 293T cells by the calcium precipitation method. Viral supernatants were concentrated by ultracentrifugation and used to transduce human LSECs.

To determine the expression of Id1, CXCR4 and CXCR7 in LSEC after cytokine stimulation, 500,000 LSECs were seeded and treated with *Cxcr4*, *Cxcr7* or scrambled shRNA lentiviruses, respectively. After starving in serum-free medium, seeded LSECs were stimulated with 10 ng ml<sup>-1</sup> SDF-1 or 20 ng ml<sup>-1</sup> FGF-2. At various time points, cells were collected for the measurement of Id1 protein and mRNA expression. Treatment with 30 µM U0126 was used to inhibit the activity of MAPK. Activation of MAPK (p-Erk1/2) was assayed by immunoblot using antibodies against p-Erk1/2 and total Erk1/2 (Cell Signaling Technology)<sup>11</sup>.

For immunoprecipitation-western blot, cell lysates were retrieved by RIPA lysis buffer with protease inhibitor cocktail and phosphatase inhibitor (Pierce) and incubated with anti-CXCR7 antibody (R&D Systems) conjugated with Protein

A/G beads (Invitrogen). Beads were retrieved by magnet, associated proteins were eluted, and the association of β-arrestin, CXCR4, and CXCR7 was determined by western blot (Santa Cruz), after normalization to total CXCR7 protein amounts in cell lysates (input).

**Flow cytometric analysis of liver non-parenchymal cells and LSECs.** For flow cytometry analysis, retrieved livers from animals were minced, dissociated in liver digestion medium (Invitrogen), and filtered through a 30-µm strainer. Single-cell suspensions were preblocked with Fc block (CD16/CD32; BD Biosciences) and then incubated with the following primary antibodies recognizing mouse LSECs and haematopoietic cells; rat IgG2ak and IgG2aβ isotype control; CD31/PECAM-1 (clone MEC 13.3, eBioscience); VE-cadherin/CD144 (clone Bv13, eBioscience); CXCR7 (clone 11G8, R&D Systems). Primary antibodies were directly conjugated to various Alexa Fluor dyes or Quantum Dots using antibody labelling kits (Invitrogen) performed as per the manufacturer's instructions. In the case of Alexa Fluor 750, conjugations were performed using succinimidyl esters and purified over BioSpin P30 Gel (Bio-Rad).

Labelled cell populations were measured by a LSRII flow cytometer (Becton Dickinson); compensation for multivariate experiments was performed with FACS Diva software. Flow cytometry analysis was performed using a variety of controls such as isotype antibodies, and unstained samples for determining appropriate gates, voltages, and compensations required in multivariate flow cytometry.

**Immunostaining and histological analysis of liver cryosection.** To collect tissues for histological analysis, mice were perfused with 4% PFA, cryopreserved, and snap frozen in OCT. For immunofluorescent microscopy, the liver sections (10 µm) were blocked (5% donkey serum/0.3% Triton X-100) and incubated in primary Abs: anti-VE-cadherin polyclonal Ab (pAb, 2 µg ml<sup>-1</sup>, R&D Systems), anti-CD31 mAb (MEC13.3, 5 µg ml<sup>-1</sup>, BD Biosciences), anti-CXCR7 mAb (5 µg ml<sup>-1</sup>, R&D Systems), anti-desmin (pAb, 2 µg ml<sup>-1</sup>, Abcam) and anti-p-Erk1/2 antibody (2 µg ml<sup>-1</sup>, Cell Signaling Technology). After incubation in fluorophore-conjugated secondary antibodies (2.5 µg ml<sup>-1</sup>, Jackson ImmunoResearch), sections were counterstained with TOPRO3 or DAPI (Invitrogen).

Liver cell proliferation *in vivo* was measured by BrdU uptake. Briefly, mice received a single dose of BrdU (Sigma) intraperitoneally 60 min before death (50 mg kg<sup>-1</sup>). Liver lobes were removed, weighed and further processed. Slices were preincubated with 1 M HCl at room temperature for 1 h, neutralized with 10 mM Tris (pH 8.5) at room temperature for 15 min, and stained using the BrdU Detection System (BD Biosciences) and fluorophore-conjugated secondary antibodies (2.5 µg ml<sup>-1</sup>, Jackson ImmunoResearch). For immunohistochemical (IHC) detection of BrdU, endogenous peroxidase and nonspecific protein block (5% BSA, 10% donkey serum and 0.02% Tween-20) were performed on liver cryosections and incubated with secondary pAb and streptavidin horseradish peroxidase (Jackson ImmunoResearch).

**Image acquisition and analysis.** IHC staining of liver slides was captured with Olympus BX51 microscope (Olympus America), and fluorescent images were recorded on AxioVert LSM710 confocal microscope (Zeiss). Co-staining of VE-cadherin with CXCR4 and CXCR7 was also determined.

**Gene expression analysis by real-time polymerase chain reaction.** Total RNA was extracted from cryopreserved liver tissue or isolated LSECs using RNeasykit (Qiagen). After isolation, 500 ng of total RNA was transcribed into complementary DNA by using the superscript reverse transcriptase Kit (Invitrogen). The detection of complementary DNA expression for the specific genes was performed using quantitative polymerase chain reaction.

31. Wang, Y. *et al.* Ephrin-B2 controls VEGF-induced angiogenesis and lymphangiogenesis. *Nature* **465**, 483–486 (2010).
32. Nie, Y. *et al.* The role of CXCR4 in maintaining peripheral B cell compartments and humoral immunity. *J. Exp. Med.* **200**, 1145–1156 (2004).



# Antibacterial membrane attack by a pore-forming intestinal C-type lectin

Sohini Mukherjee<sup>1</sup>, Hui Zheng<sup>2</sup>, Mehabaw G. Derebe<sup>1</sup>, Keith M. Callenberg<sup>3</sup>, Carrie L. Partch<sup>4</sup>, Darcy Rollins<sup>1</sup>, Daniel C. Propheter<sup>1</sup>, Josep Rizo<sup>5</sup>, Michael Grabe<sup>3†</sup>, Qiu-Xing Jiang<sup>2\*</sup> & Lora V. Hooper<sup>1,6\*</sup>

**Human body-surface epithelia coexist in close association with complex bacterial communities and are protected by a variety of antibacterial proteins. C-type lectins of the RegIII family are bactericidal proteins that limit direct contact between bacteria and the intestinal epithelium and thus promote tolerance to the intestinal microbiota<sup>1,2</sup>. RegIII lectins recognize their bacterial targets by binding peptidoglycan carbohydrate<sup>1,3</sup>, but the mechanism by which they kill bacteria is unknown. Here we elucidate the mechanistic basis for RegIII bactericidal activity. We show that human RegIII $\alpha$  (also known as HIP/PAP) binds membrane phospholipids and kills bacteria by forming a hexameric membrane-permeabilizing oligomeric pore. We derive a three-dimensional model of the RegIII $\alpha$  pore by docking the RegIII $\alpha$  crystal structure into a cryo-electron microscopic map of the pore complex, and show that the model accords with experimentally determined properties of the pore. Lipopolysaccharide inhibits RegIII $\alpha$  pore-forming activity, explaining why RegIII $\alpha$  is bactericidal for Gram-positive but not Gram-negative bacteria. Our findings identify C-type lectins as mediators of membrane attack in the mucosal immune system, and provide detailed insight into an antibacterial mechanism that promotes mutualism with the resident microbiota.**

RegIII $\alpha$  damages the surfaces of Gram-positive bacteria<sup>1</sup>, indicating that RegIII $\alpha$  might target bacterial membranes. We assessed the capacity of RegIII $\alpha$  to permeabilize bacterial membranes by quantifying bacterial uptake of a membrane-impermeant fluorescent dye (SYTOX green). RegIII $\alpha$  increased SYTOX green uptake when added to the Gram-positive species *Listeria monocytogenes*, indicating damaged membranes (Fig. 1a, b). RegIII $\alpha$  has an anionic amino-terminal pro-segment that inhibits bactericidal activity (but not peptidoglycan binding) by docking to the protein core through charge–charge interactions<sup>4</sup>. The pro-segment is removed by trypsin on secretion into the intestinal lumen, yielding bactericidally active RegIII $\alpha$  (ref. 4). Bactericidally inactive pro-RegIII $\alpha$  did not induce SYTOX green uptake, indicating minimal membrane permeabilization (Fig. 1a). Thus, RegIII $\alpha$  permeabilizes the bacterial membrane, and the pro-segment inhibits this activity.

To test directly for membrane disruption by RegIII $\alpha$  we used liposomes composed of 85% zwitterionic phospholipid (PC) and 15% acidic phospholipid (PS). The liposomes encapsulated carboxyfluorescein, a fluorescent dye. RegIII $\alpha$  induced rapid dye efflux from PC/PS liposomes (Fig. 1c), which was reduced when PC-only liposomes were used (Fig. 1d, e). This indicates a preference for acidic phospholipids that is consistent with the acidic lipid content of bacterial membranes<sup>5</sup> and with the salt sensitivity of RegIII $\alpha$  membrane toxicity (Extended Data Fig. 2a, b). These findings indicate that RegIII $\alpha$  interactions with lipid bilayers are mediated by electrostatic interactions. pro-RegIII $\alpha$  yielded a diminished rate of dye release (Fig. 1f), indicating that the pro-segment inhibits membrane permeabilization.

We next assessed RegIII $\alpha$  lipid-binding activity by measuring changes in the intrinsic fluorescence of tryptophan residues<sup>6</sup>. We observed increased tryptophan fluorescence intensity only when RegIII $\alpha$  was added to PS-containing liposomes (Fig. 1g–i), indicating that RegIII $\alpha$  interacts with acidic phospholipids. Furthermore, we observed fluorescence resonance energy transfer (FRET) between donor RegIII $\alpha$  tryptophan residues and dansyl-labelled PC/PS liposomes<sup>7</sup> (Fig. 1j, k). FRET was inhibited by the pro-RegIII $\alpha$  N-terminal pro-segment (Fig. 1j, k), indicating that the pro-segment inhibits bactericidal activity by hindering lipid binding. Consistent with its inability to bind lipids, pro-RegIII $\alpha$  did not inhibit RegIII $\alpha$  bactericidal activity in mixing experiments (Extended Data Fig. 2c).

Several membrane-active toxins destabilize membranes by forming monomeric or multimeric pores<sup>8</sup>. To test for RegIII $\alpha$  pores, we performed conductance studies in black lipid membranes, a model system that mimics the properties of a cell membrane<sup>9</sup>. RegIII $\alpha$  produced rapid single-channel-like currents at  $-80$  mV in the presence of  $Mg^{2+}$  ions (Fig. 2a), with no current detected at 0 mV. Using the Nernst–Planck equation we estimated the diameter of the pore at  $\sim 12$ – $14$  Å (Extended Data Fig. 3). The calculated pore size agreed with the lack of efflux of fluorescein isothiocyanate-dextran-10 (FD10) or FD4, which have Stokes diameters of  $\sim 44$  Å and  $\sim 28$  Å, respectively (Fig. 2b). In contrast, carboxyfluorescein ( $\sim 10$  Å) passed readily through the pores (Figs 1c and 2b). These results show that RegIII $\alpha$  forms functional transmembrane pores and yield an estimate of the inner pore diameter.

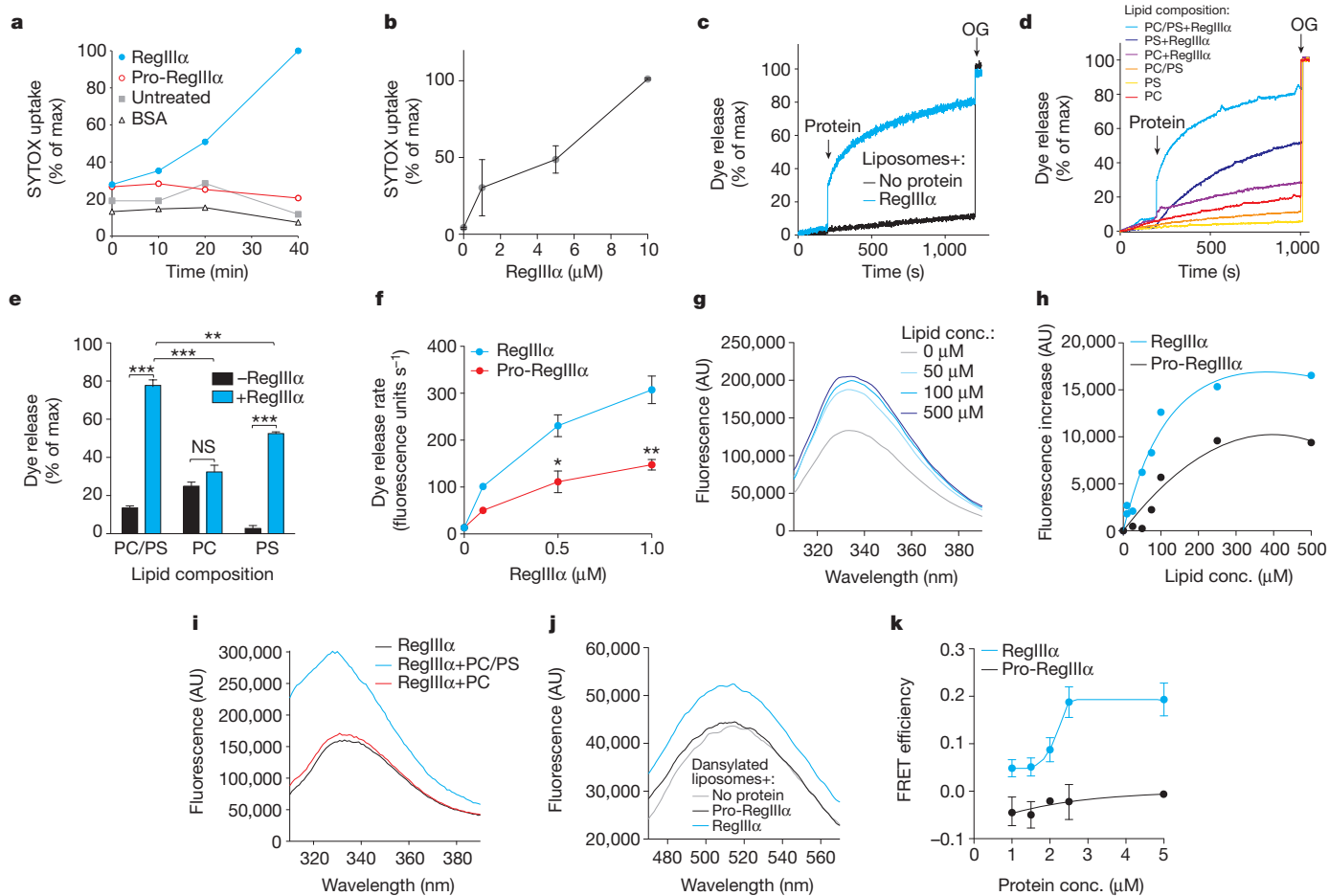
When visualized by negative-stain electron microscopy (EM), numerous circular structures of  $\sim 100$  Å diameter were observed on liposomes incubated with RegIII $\alpha$  (Fig. 2c and Extended Data Fig. 4a). Although RegIII $\alpha$  is a monomer in solution<sup>10</sup>, the size of the pores suggested that they were RegIII $\alpha$  multimers. We therefore treated liposome-associated RegIII $\alpha$  with a crosslinking agent, solubilized the products in detergent, and separated them by size-exclusion chromatography (Fig. 2d). In addition to a prominent monomer peak we detected a second, liposome-dependent peak at a lower retention volume, indicating the formation of a multimeric complex. Western blotting showed a single RegIII $\alpha$  species with mobility similar to that predicted for a hexamer (Fig. 2d), suggesting that the pore was a RegIII $\alpha$  hexamer.

After longer incubations with lipid, RegIII $\alpha$  formed filaments (Extended Data Fig. 4b) similar to those in pancreatic secretions<sup>11</sup>. The filaments were  $\sim 100$  Å in diameter, correlating with the dimensions of the RegIII $\alpha$  pore (Fig. 2c). RegIII $\alpha$  filamentation required lipid and was dependent on RegIII $\alpha$  pore formation, as pro-RegIII $\alpha$  formed neither pores nor filaments (Extended Data Fig. 4b, d). Filamentation partially inhibited the ability of RegIII $\alpha$  to permeabilize membranes (Extended Data Figs 4c and 5a–c), as observed with other membrane toxic host defence proteins where filamentation traps pore complexes and limits damage to host cells<sup>12</sup>. These findings indicate that the RegIII $\alpha$  filaments

<sup>1</sup>Department of Immunology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. <sup>2</sup>Department of Cell Biology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. <sup>3</sup>Department of Biological Sciences, University of Pittsburgh, and Joint Carnegie Mellon University–University of Pittsburgh PhD Program in Computational Biology, Pittsburgh, Pennsylvania 15261, USA. <sup>4</sup>Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95064, USA. <sup>5</sup>Department of Biochemistry and Department of Pharmacology, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. <sup>6</sup>The Howard Hughes Medical Institute, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.

<sup>†</sup>Present address: Cardiovascular Research Institute, University of California, San Francisco, San Francisco, California 94143, USA.

\*These authors contributed equally to this work.



**Figure 1 | RegIIIα permeabilizes the bacterial membrane.** **a**, *Listeria monocytogenes* was treated with 25 μM RegIIIα, pro-RegIIIα, or BSA or left untreated, and bacterial uptake of SYTOX green was measured. Results are representative of three independent experiments, and are expressed as a percentage of maximum SYTOX uptake in the presence of 0.2% SDS. **b**, SYTOX green uptake by *L. monocytogenes* in the presence of increasing RegIIIα concentrations. Assays were performed in triplicate. Means ± s.e.m. are plotted. **c**, Carboxyfluorescein (CF)-loaded liposomes (10 μM lipid; 85% PC/15% PS) were treated with 1 μM RegIIIα. 1.0% octylglucoside (OG) was added towards the end to disrupt remaining liposomes. Dye efflux is expressed as percentage of maximal release by detergent. Results are representative of five independent experiments. **d**, 10 μM RegIIIα was added to carboxyfluorescein-loaded liposomes (100 μM lipid; 100% PC, 100% PS or 85% PC:15% PS), and dye efflux was monitored over time. Representative results are shown.

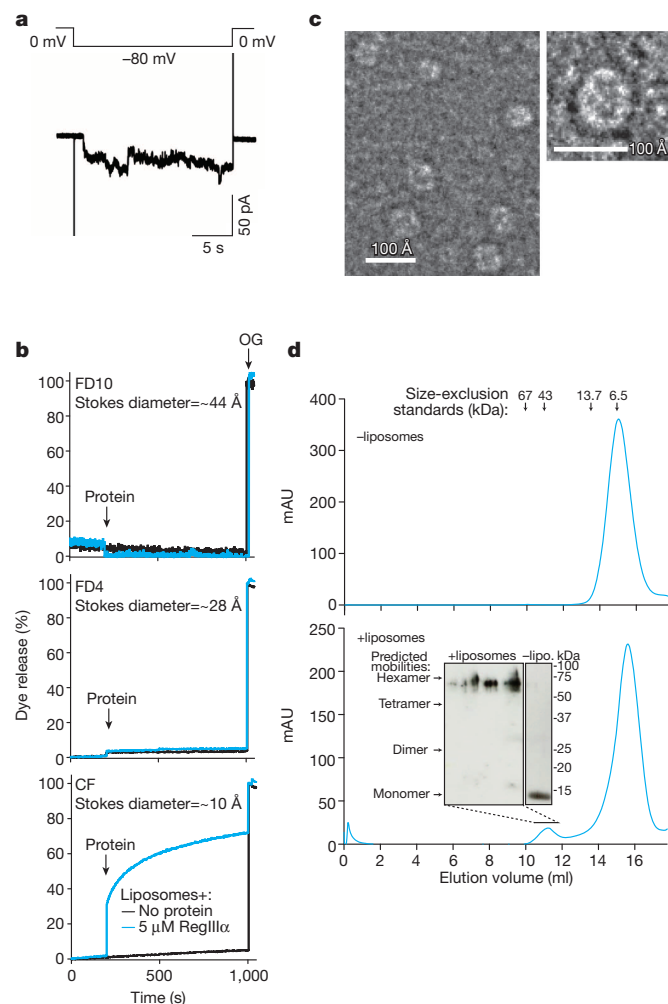
are higher-order assemblies of RegIIIα pore complexes and show that filamentation limits RegIIIα toxicity.

Although the ~90-kDa RegIIIα pore complex was too small for structure determination by single-particle cryoelectron microscopy (cryoEM) methods<sup>13</sup>, the RegIIIα filaments were sufficiently large for such analysis. We therefore reconstructed a three-dimensional map of the RegIIIα filament and extracted the structure of the minimal pore complex (Fig. 3a, b and Extended Data Fig. 6a–f). The nominal resolution of our structure, 9.2 Å, was limited by symmetry variability and filament bending (Extended Data Fig. 6g–j and Supplementary Information). Consistent with our crosslinking studies (Fig. 2d), the minimal pore was a hexamer formed by three RegIIIα dimers related by helical symmetry. The outer diameter of the pore assembly was 89 Å, as observed by negative-stain EM (Fig. 2c). The pore height was 55 Å, sufficient to span a lipid bilayer (35–45 Å)<sup>14</sup>. The inner diameter was ~18 Å, consistent with the pore size predicted by our conductance measurements (Extended Data Fig. 3) and dye release assays (Fig. 2b).

**e**, Averaged results from three independent replicates of the experiment shown in **d**. NS, not significant; \*\**P* < 0.01; \*\*\**P* < 0.001. **f**, Initial rate of liposome dye efflux (100 μM lipid) as a function of RegIIIα and pro-RegIIIα concentration. Results are representative of three independent experiments. \**P* < 0.05; \*\**P* < 0.01. **g**, Intrinsic tryptophan fluorescence of 1 μM RegIIIα was measured in the presence of increasing lipid concentrations. **h**, Tryptophan fluorescence of 1 μM RegIIIα and pro-RegIIIα as a function of lipid concentration. **i**, Intrinsic tryptophan fluorescence of 1 μM RegIIIα was measured in the presence of liposomes (100 μM lipid) of varying lipid composition. **j**, 5.0 μM RegIIIα or pro-RegIIIα was added to liposomes (100 μM lipid) incorporating 5% dansyl-PE and dansyl fluorescence was monitored. Assays were performed in triplicate. **k**, FRET efficiency as a function of RegIIIα and pro-RegIIIα concentration. Assays were performed in triplicate. Means ± s.e.m. are plotted.

RegIIIα, like other epithelial bactericidal proteins such as α-defensins, is constrained by disulphide bonds that prohibit large secondary structure changes on moving from an aqueous to an apolar environment<sup>15,16</sup>. This suggested the feasibility of docking the three-dimensional structure of the RegIIIα monomer into the EM density map to model the organization of the pore complex further. First, we determined the crystal structure of processed, bactericidally active RegIIIα (Extended Data Fig. 7a) and compared it to the previously determined structure of bactericidally inactive pro-RegIIIα. The two structures were similar, although the amino acid side chains of the loop encompassing residues 93–99 (sequence KSIGNSY) adopted different orientations in the active RegIIIα structure (Fig. 3c). This was consistent with the conformational flexibility of this loop as indicated by a higher crystallographic *B*-factor (Extended Data Fig. 7b).

The active RegIIIα structure could be docked into the cryo-EM hexameric density map (Fig. 3d and Extended Data Fig. 6k, l), providing good spatial constraints for building a hexameric model. The model indicates that the RegIIIα subunits in the pore assembly are



**Figure 2 | RegIII $\alpha$  forms a transmembrane pore.** **a**, RegIII $\alpha$ -dependent current flow across a planar lipid bilayer is depicted as a function of time. No current was observed before the application of a voltage across the membrane. Upon the application of  $-80$  mV, inward current was observed, and returning the membrane potential to zero diminished the current because the measured reverse potential was  $-4.0$  mV. The current trace is representative of multiple independent experiments. **b**, Liposomes loaded with FITC-Dextran 10 (FD10), FITC-Dextran 4 (FD4), or carboxyfluorescein (CF) were treated with  $5.0$   $\mu$ M RegIII $\alpha$  and dye release was monitored over time.  $1\%$  OG (octylglucoside) was added to disrupt the liposomes towards the end of the experiment. **c**, Negative-stain electron microscopy (EM) images of RegIII $\alpha$  in the presence of lipid bilayers. An individual RegIII $\alpha$  pore particle is shown in the right-hand panel. **d**, RegIII $\alpha$  ( $100$   $\mu$ M) in the presence or absence of liposomes ( $1$  mM lipid) was crosslinked with  $5$  mM 1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride (EDC). Crosslinked complexes were solubilized in detergent, resolved by size-exclusion chromatography, and analysed by western blotting with anti-RegIII antibody. The predicted mobilities of RegIII $\alpha$  dimers, tetramers and hexamers were calculated from the mobility of the monomer after crosslinking in the absence of liposomes (right panel in blot).

oriented with the carbohydrate-binding loop pointing towards the central channel, and the loop encompassing residues 93–99 and the N and carboxy (C) termini oriented towards the lipid bilayer (Fig. 3d). The resolution of our map did not allow us to extract detailed information about intermolecular interactions in the pore complex. There was imperfect docking of the carbohydrate-binding loop, the loop encompassing residues 93–99, and the far N terminus (Fig. 3d), consistent with the conformational flexibility of these regions (Extended Data Fig. 7b).

We used mutagenesis to assess experimentally the orientation of RegIII $\alpha$  in the pore complex. Our model predicts that the basic residue

Lys 93 is oriented towards the lipid bilayer (Fig. 3d) and thus might be involved in interactions with the negatively charged phospholipids required for RegIII $\alpha$ –liposome interactions (Fig. 1d, e). A Lys93Ala mutation, but not conservative Lys93Arg and Lys93His mutations, reduced the toxicity of RegIII $\alpha$  for liposomes as well as intact bacteria (Fig. 3e, f and Extended Data Fig. 8a). In contrast, a Glu114Gln mutation, which resides in the carbohydrate-binding loop (Fig. 3c)<sup>3</sup>, did not have an impact on membrane toxicity, consistent with its predicted position near the pore interior (Fig. 3d, e). As expected, the Lys93Ala mutation but not the Glu114Gln mutation inhibited filament formation (Extended Data Fig. 8b). Finally, the orientation of the N terminus towards the lipid bilayer is consistent with the role of the N-terminal pro-segment in inhibiting RegIII $\alpha$  interactions with lipid and reducing membrane toxicity (Fig. 1a, f, h, j, k).

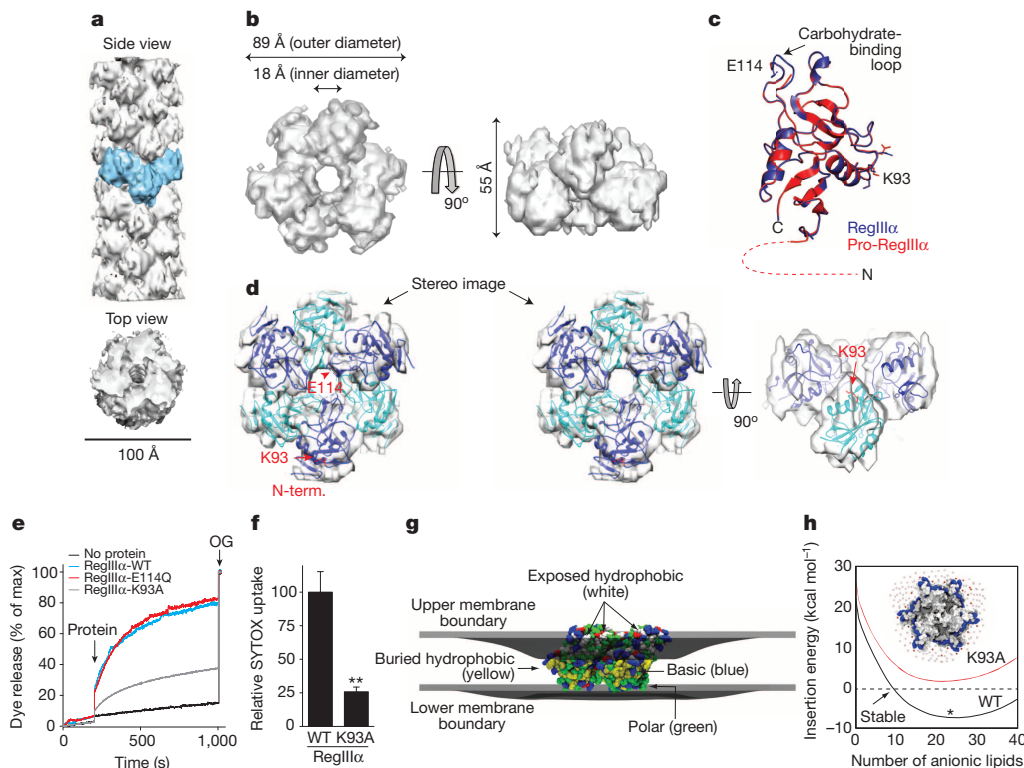
We next calculated the energetics of pore insertion into a PC-like membrane bilayer using physics-based computational modelling (Extended Data Fig. 9a–d)<sup>17</sup>. The model predicts that basic residues are located near the membrane–water interface whereas a strip of hydrophobic and polar residues is buried in the membrane core (Fig. 3g). The complex presents a positive electric field to the membrane (Extended Data Fig. 9e, f), creating an unfavourable electrostatic energy unless negatively charged PS-like lipids are added to the membrane (Fig. 3h). This is consistent with our finding that PS lipids are necessary for RegIII $\alpha$  toxicity (Fig. 1d, e). Finally, calculations on the Lys93Ala mutant showed reduced stability (Fig. 3h) due to loss of favourable electrostatic interactions between Lys 93 and negatively charged lipids. Thus, the model reveals that charge sequestration is a critical determinant of RegIII $\alpha$  pore stability in the membrane. Furthermore, the model predicts that Arg 166 interacts with the membrane surface (Extended Data Fig. 10a). Consistent with this prediction, an Arg166Ala mutation reduced membrane toxicity of RegIII $\alpha$  (Extended Data Fig. 10b). In contrast, mutating Arg 39, which is exposed to aqueous solvent in the model, had little effect on RegIII $\alpha$  membrane toxicity (Extended Data Fig. 10a, b). Thus, our model accurately predicts the experimental behaviour of the RegIII $\alpha$  pore.

RegIII $\alpha$  selectively targets Gram-positive bacteria<sup>1</sup>, raising the question of why RegIII $\alpha$  cannot kill Gram-negative bacteria by permeabilizing the outer membrane. In contrast to PC/PS liposomes, liposomes composed of an *Escherichia coli* total lipid extract were not disrupted by RegIII $\alpha$  (Fig. 4a), indicating that a component of the lipid extract inhibited membrane permeabilization. Lipopolysaccharide (LPS), a major constituent of the Gram-negative outer membrane, inhibited RegIII $\alpha$ -mediated liposome disruption and antibacterial activity (Fig. 4b, c), indicating that LPS is one factor that prevents RegIII $\alpha$ -mediated permeabilization of Gram-negative bacteria.

Finally, we postulated that the trypsin-cleavable inhibitory N terminus of pro-RegIII $\alpha$  evolved to suppress pore-forming activity and thus minimize cytotoxicity during RegIII $\alpha$  synthesis and storage in epithelial cells. In support of this idea, RegIII $\alpha$  was cytotoxic towards cultured intestinal epithelial cells (MODE-K)<sup>18</sup>, and the pro-segment suppressed this cytotoxicity (Fig. 4d, e).

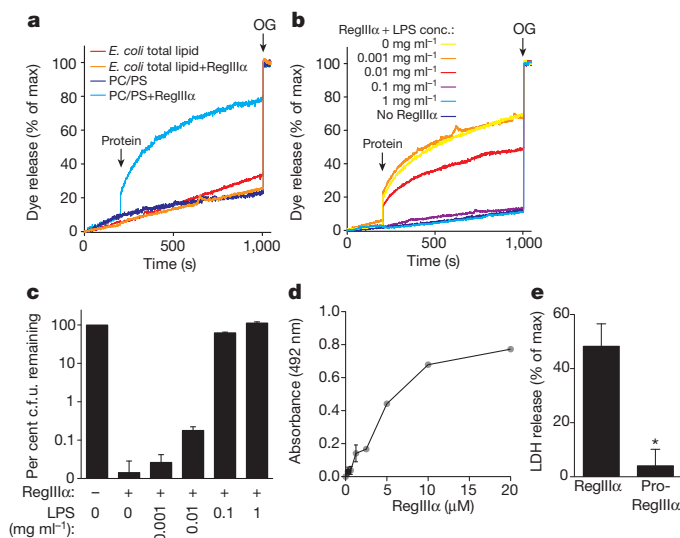
Thus, RegIII $\alpha$  kills its bacterial targets by oligomerizing on the bacterial membrane to form a membrane-penetrating pore (Extended Data Fig. 1). Membrane attack by pore formation represents a previously unappreciated biological activity for the C-type lectin family. Our findings may provide insight into the evolutionary origins of the lectin-mediated complement pathway, in which recruited complement proteins disrupt microbial membranes<sup>19</sup>. With its intrinsic capacity for membrane attack, RegIII $\alpha$  may represent a more evolutionarily primitive mechanism of lectin-mediated innate immunity. We propose that the lectin-mediated complement pathway could have evolved from a directly bactericidal ancestral lectin, with the bacterial recognition function retained by the descendent C-type lectin(s) and the membrane attack function assumed by recruited accessory proteins that assemble into the membrane attack complex.





**Figure 3 | Structural model of the RegIIIα pore complex.** **a**, Top and side view of the cryoEM reconstruction of the RegIIIα filament. **b**, Top and side view of the cryoEM map of the RegIIIα hexameric complex at a nominal 9.2 Å resolution. **c**, Ribbon representation of the crystal structure of active monomeric RegIIIα (Protein Data Bank (PDB) code 4MTH), aligned with the pro-RegIIIα structure (PDB code 1UV0). The first ten residues of the N-terminal pro-segment are disordered and are therefore missing from the structure; these residues have been depicted as a dashed red line. Side chains in the loop encompassing amino acids 93–99 (KSIGNSY) are shown as sticks. **d**, Stereo diagram showing docking of the active RegIIIα crystal structure into the cryoEM density map. The docked structures are alternately coloured blue and cyan to aid in visualization of the individual subunits. The positions of Lys 93 (K93) and Glu 114 (E114) are indicated. **e**, 5 μM of wild-type (WT), Lys93Ala (K93A) mutant, or Glu114Gln (E114Q) mutant RegIIIα was added to 100 μM carboxyfluorescein-loaded liposomes and dye efflux was monitored. **f**, 1 μM wild-type or Lys93Ala mutant RegIIIα was assayed for membrane disruption in bacteria using the SYTOX uptake assay described in Fig. 1. Assays were performed in triplicate and results are expressed relative to wild-type

RegIIIα. Error bars indicate s.e.m.; \*\* $P < 0.01$ . **g**, Most energetically stable membrane configuration around the embedded hexamer. The upper membrane boundary (grey surface) bends down to expose large charged portions of the protein to water, whereas the lower membrane boundary (grey surface) exhibits minor deflections. The region between the upper and lower boundaries is a water-inaccessible region composed of the high-dielectric head-groups and the low-dielectric core. A stretch of hydrophobic residues (yellow) is in the centre of the membrane, whereas charged (basic in blue and acidic in red) and polar (green) residues are near the upper and lower membrane boundaries in the high-dielectric head-group region. **h**, Using the configuration in **g**, we added negatively charged point charges to the head-group regions to model addition of PS lipids (red dots in the inset model). At low values, the total insertion energy for the wild-type protein is positive, indicating a lack of stability, but above 10 negatively charged lipids, the hexamer is stabilized in the membrane (black curve). The optimal lipid configuration is indicated by an asterisk. The insertion energy for the Lys93Ala mutant is in red. Inset: top-down view; red dots, PS lipids; blue, Arg and Lys residues; white dots, uncharged lipid positions.



**Figure 4 | Regulation of RegIIIα pore formation.** **a–c**, RegIIIα pore formation is inhibited by lipopolysaccharide. **a**, 10 μM RegIIIα was added to liposomes composed of lipids from an *E. coli* total lipid extract or from PC/PS as a control. **b**, 10 μM RegIIIα was added to liposomes (100 μM lipid) in the presence of varying LPS concentrations. **c**, 10 μM RegIIIα was added to  $\sim 10^4$  c.f.u. of log phase *L. monocytogenes* in the presence of varying LPS concentrations. The assay was carried out at 37 °C for 2 h, and surviving bacteria were quantified by dilution plating. Assays were done in triplicate. Results in **a–c** are representative of two independent experiments. **d, e**, The RegIIIα N-terminal pro-segment limits toxicity towards mammalian cells. **d**, RegIIIα was added to MODE-K cells and cytotoxicity was determined by quantifying lactate dehydrogenase (LDH) release. LDH activity was assessed by spectrophotometric detection of an enzymatic product of LDH at 492 nm. **e**, 10 μM RegIIIα or pro-RegIIIα was added to MODE-K cells and LDH release was quantified. Maximum LDH release was determined by treating cells with NP-40 detergent.



## METHODS SUMMARY

**Preparation of recombinant RegIII $\alpha$ .** Recombinant human pro-RegIII $\alpha$  and RegIII $\gamma$  were expressed and purified according to published methods<sup>1,4</sup>.

**Membrane permeabilization assays.** *Listeria monocytogenes* was exposed to 25  $\mu$ M RegIII $\alpha$ , pro-RegIII $\alpha$  or bovine serum albumin (BSA), incubated with SYTOX green, and cell-associated fluorescence was quantified. For dye leakage assays, fluorescence of carboxyfluorescein-loaded liposomes was monitored over time on a PTI spectrofluorometer, in the presence or absence of RegIII $\alpha$  or pro-RegIII $\alpha$ .

**Lipid binding assays.** Binding of RegIII $\alpha$  and pro-RegIII $\alpha$  to liposomes was measured by monitoring fluorescence resonance energy transfer (FRET) between protein tryptophan residues and dansyl-PE. Fluorescence spectra were recorded on a PTI Spectrofluorometer. Measurements of intrinsic tryptophan fluorescence of RegIII $\alpha$  in the absence or presence of liposomes were recorded on a PTI Spectrofluorometer between 290 and 450 nm at a fixed excitation wavelength of 280 nm.

**Crosslinking experiments.** RegIII $\alpha$  was incubated with liposomes for 20 min followed by 1 h treatment with 5 mM of the crosslinking reagent, EDC, at room temperature. The samples were solubilized with 40 mM *n*-decyl- $\beta$ -D-maltopyranoside (DM) detergent, separated by size-exclusion chromatography, and analysed by western blotting with detection by anti-RegIII $\gamma$  antibody<sup>10</sup>.

**Determination of the RegIII $\alpha$  crystal structure.** Recombinant RegIII $\alpha$  lacking the N-terminal pro-segment was crystallized using the sitting-drop vapour diffusion method. We collected X-ray diffraction data at the Advanced Photon Source, Argonne National Laboratory. The structure was determined by molecular replacement using a starting model of the full-length RegIII $\alpha$  structure, followed by cycles of model building. Further details are available in Supplementary Information.

**CryoEM imaging.** Images were acquired in a JEOL JEM2200FS FEG transmission electron microscope equipped with an in-column energy filter. Full details are available in Supplementary Information.

**Computational modelling studies.** Full details are available in Supplementary Information.

**Statistical analysis.** All *P* values were calculated using the unpaired, two-tailed *t*-test.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 February; accepted 2 October 2013.

Published online 20 November 2013.

- Cash, H. L., Whitham, C. V., Behrendt, C. L. & Hooper, L. V. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* **313**, 1126–1130 (2006).
- Vaishnav, S. *et al.* The antibacterial lectin RegIII $\gamma$  promotes the spatial segregation of microbiota and host in the intestine. *Science* **334**, 255–258 (2011).
- Lehotzky, R. E. *et al.* Molecular basis for peptidoglycan recognition by a bactericidal lectin. *Proc. Natl Acad. Sci. USA* **107**, 7722–7727 (2010).
- Mukherjee, S. *et al.* Regulation of C-type lectin antimicrobial activity by a flexible N-terminal prosegment. *J. Biol. Chem.* **284**, 4881–4888 (2009).
- Zaslöff, M. Antimicrobial peptides of multicellular organisms. *Nature* **415**, 389–395 (2002).
- Kraft, C. A., Garrido, J. L., Leiva-Vega, L. & Romero, G. Quantitative analysis of protein-lipid interactions using tryptophan fluorescence. *Sci. Signal.* **2**, 14 (2009).

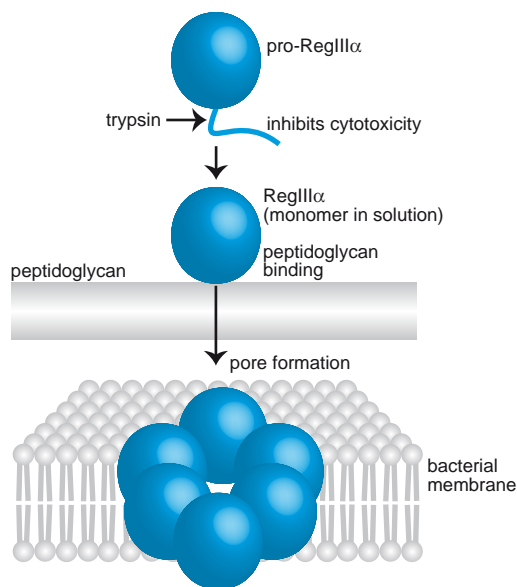
- Fernandez, I. *et al.* Three-dimensional structure of the synaptotagmin 1 C2B-domain: synaptotagmin 1 as a phospholipid binding machine. *Neuron* **32**, 1057–1069 (2001).
- Brogden, K. A. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature Rev. Microbiol.* **3**, 238–250 (2005).
- Mueller, P., Rudin, D. O., Tien, H. T. & Wescott, W. C. Reconstitution of cell membrane structure *in vitro* and its transformation into an excitable system. *Nature* **194**, 979–980 (1962).
- Cash, H. L., Whitham, C. V. & Hooper, L. V. Refolding, purification, and characterization of human and murine RegIII proteins expressed in *Escherichia coli*. *Protein Expr. Purif.* **48**, 151–159 (2006).
- Ho, M.-R. *et al.* Human pancreatitis-associated protein forms fibrillar aggregates with a native-like conformation. *J. Biol. Chem.* **281**, 33566–33576 (2006).
- Kagan, B. L. *et al.* Antimicrobial properties of amyloid peptides. *Mol. Pharm.* **9**, 708–717 (2012).
- Jiang, Q.-X., Wang, D.-N. & MacKinnon, R. Electron microscopic analysis of KvAP voltage-dependent K<sup>+</sup> channels in an open conformation. *Nature* **430**, 806–810 (2004).
- Lodish, H. *et al.* in *Molecular Cell Biology* (W.H. Freeman, 2000).
- Kagan, B. L., Selsted, M. E., Ganz, T. & Lehrer, R. I. Antimicrobial defensin peptides form voltage-dependent ion-permeable channels in planar lipid bilayer membranes. *Proc. Natl Acad. Sci. USA* **87**, 210–214 (1990).
- Zhang, Y., Lu, W. & Hong, M. The membrane-bound structure and topology of a human  $\alpha$ -defensin indicate a dimer pore mechanism for membrane disruption. *Biochemistry* **49**, 9770–9782 (2010).
- Callenberg, K. M., Latorraca, N. R. & Grabe, M. Membrane bending is critical for the stability of voltage sensor segments in the membrane. *J. Gen. Physiol.* **140**, 55–68 (2012).
- Vidal, K., Grosjean, I., Revillard, J.-P., Gespach, C. & Kaiserlian, D. Immortalization of mouse intestinal epithelial cells by the SV40-large T gene. Phenotypic and immune characterization of the MODE-K cell line. *J. Immunol. Methods* **166**, 63–73 (1993).
- Ip, W. K. E., Takahashi, K., Ezekowitz, R. A. & Stuart, L. M. Mannose-binding lectin and innate immunity. *Immunol. Rev.* **230**, 9–21 (2009).

**Supplementary Information** is available in the online version of the paper.

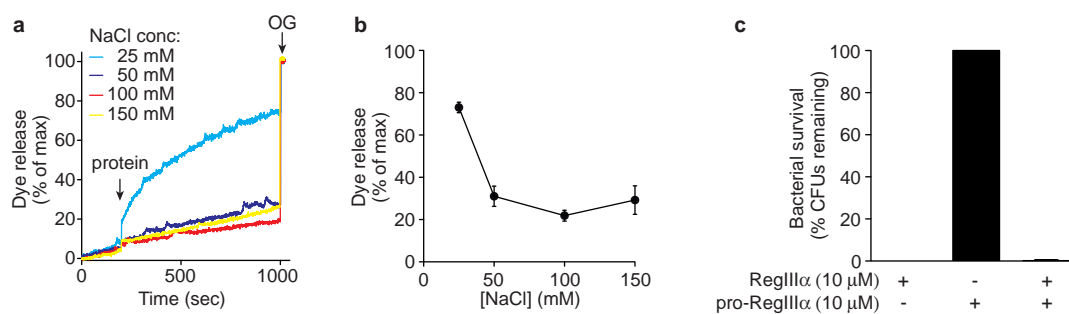
**Acknowledgements** We thank T. Craig for assistance with the liposome disruption assays. The MODE-K cell line was provided by D. Kaiserlian, INSERM U851, Lyon, France. We thank E. Egelman for sharing his programs and for offering advice on cryoEM data analysis. This work was supported by NIH R01 DK070855 (L.V.H.), NIH R01 GM088745 and GM093271 (Q.-X.J.), NIH R01 NS049444 (J.R.), Welch Foundation (I-1684 to Q.-X.J.), NSF CAREER MCB0845286 (M.G.), a Helen Hay Whitney Fellowship (S.M.), a Burroughs Wellcome Foundation New Investigators in the Pathogenesis of Infectious Diseases Award (L.V.H.), and the Howard Hughes Medical Institute (L.V.H.). Part of this work was performed in laboratories constructed with support from NIH grant C06 RR30414.

**Author Contributions** S.M., M.G., Q.-X.J. and L.V.H. designed the research, analysed data, and wrote the paper. S.M., H.Z., C.L.P., D.R. and D.C.P. performed most of the experiments. M.G.D. determined the crystal structure of bactericidally active human RegIII $\alpha$ . H.Z. performed the bilayer recordings. K.M.C. and M.G. performed the physics-based computational modelling studies. S.M., H.Z., C.L.P., J.R. M.G., Q.-X.J. and L.V.H. interpreted the data.

**Author Information** Coordinates of the crystallographic structure of active human RegIII $\alpha$  have been deposited in the Protein Data Bank with accession code 4MTH. The cryoEM map has also been deposited in the 3D EM database under accession code EMD-5795. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.V.H. ([lora.hooper@utsouthwestern.edu](mailto:lora.hooper@utsouthwestern.edu)) or Q.-X.J. ([qiu-xing.jiang@utsouthwestern.edu](mailto:qiu-xing.jiang@utsouthwestern.edu)).

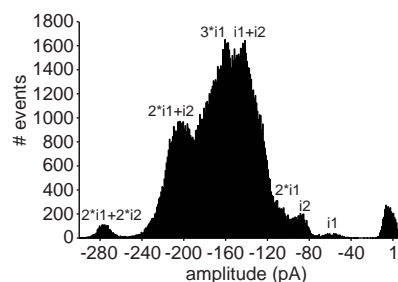


**Extended Data Figure 1 | Model of RegIIIα bactericidal function.** An overall model that incorporates both the peptidoglycan and lipid-binding functions of RegIIIα is depicted. Combining our current and previous findings, we propose that RegIIIα recognizes and kills its bacterial targets in two distinct steps. First, RegIIIα is secreted from epithelial cells as a soluble monomer that recognizes Gram-positive bacteria by binding to peptidoglycan carbohydrate via an EPN motif located in the long loop region<sup>1,3</sup>. Second, RegIIIα kills bacteria by oligomerizing in the bacterial membrane to form a hexameric membrane-penetrating pore that is predicted to induce uncontrolled ion efflux with subsequent osmotic lysis. The inhibitory N terminus of pro-RegIIIα hinders lipid binding and consequently suppresses pore formation until it is removed by trypsin after secretion into the intestinal lumen<sup>4</sup>. We propose that the inhibitory N-terminal peptide evolved to minimize collateral damage from the RegIIIα pore-forming activity during RegIIIα storage in the membrane-bound secretory granules of epithelial cells. In support of this idea, RegIIIα damages mammalian cell membranes and the N-terminal pro-segment limits this toxicity (Fig. 4d, e).



**Extended Data Figure 2 | Characterization of RegIIIα membrane permeabilization activity.** **a, b**, Impact of NaCl concentration on RegIIIα membrane permeabilization activity. **a**, 10 μM RegIIIα was added to liposomes (100 μM lipid) in the presence of varying NaCl concentrations. Representative results are shown. **b**, Averaged results from three independent replicates of the

experiment shown in **a**. **c**, Pro-RegIIIα does not inhibit RegIIIα bactericidal activity. 10 μM of purified recombinant pro-RegIIIα, RegIIIα, or a combination of the two was added to  $\sim 10^5$  c.f.u. of *L. monocytogenes* for 2 h at 37 °C. Surviving bacteria were quantified by dilution plating.

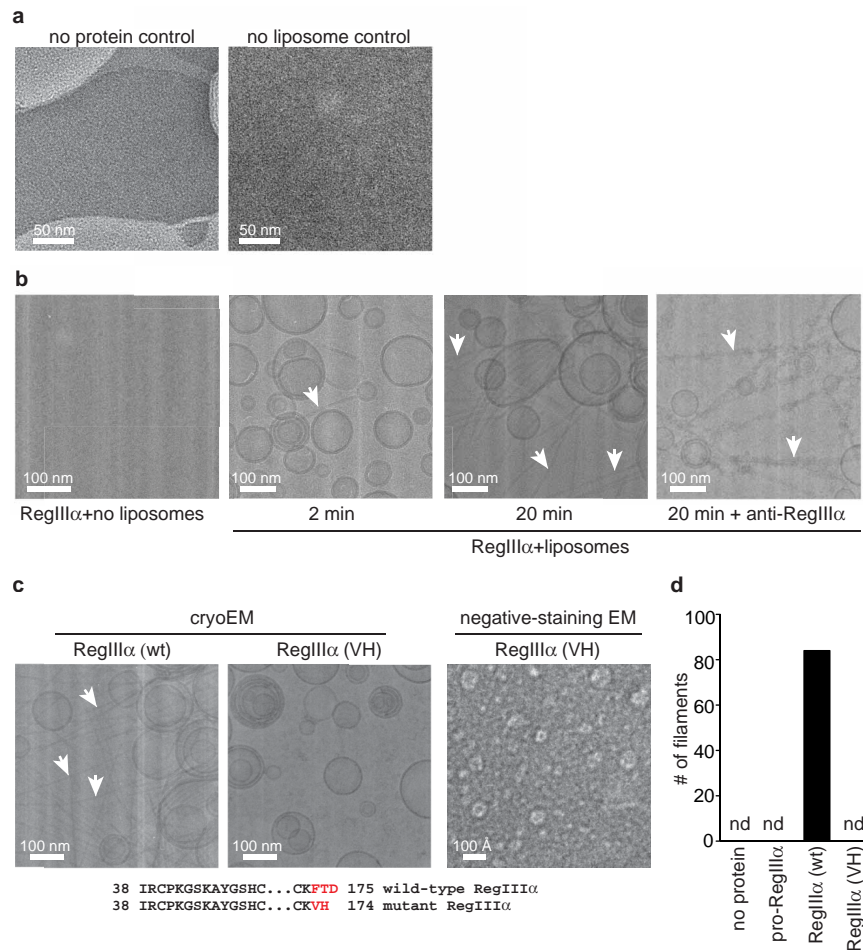


**Extended Data Figure 3 | RegIII $\alpha$  forms a transmembrane pore.** Analysis of RegIII $\alpha$  conductance in lipid bilayers. The trace of a typical single channel recording gave rise to the event histogram shown here. At  $-80$  mV, there was a short latency before the first opening event, which led to the baseline current of  $-6.5$  pA at  $-80$  mV. The baseline current was subtracted so that the baseline corresponds to a peak at  $0$  pA. Once we assigned two basic peaks at  $-53$  pA and  $-81$  pA as two independent opening events ( $i1$  and  $i2$ ), all the other major peaks in the histogram are linear combinations of these two basic events (as labelled). The data therefore suggested two different scenarios. One is that there are three pores, and each pore has two different conducting states, which may reflect the flexible diameter of the pore. The other is that  $i1$  and  $i2$  reflect two different pores that have different diameters, and that there are at least five different channels in the membrane to produce the observed histogram. This second scenario correlates with the observed variability in helical symmetry. With the idea of variability and protein dynamics in mind, it is likely that the two types of pores may interconvert with each other in the membrane. From the basic events, we estimated the pore diameters by applying the Nernst–Planck equation. In the experimental conditions, our recording chambers had  $150$  mM  $K^+$ ,  $25$  mM  $Na^+$ ,  $215$  mM  $Cl^-$ ,  $20$  mM  $Mg^{2+}$  and  $10$  mM MES pH  $5.5$  in the *cis* side, and  $20$  mM  $K^+$ ,  $25$  mM  $Na^+$ ,  $45$  mM  $Cl^-$  and  $10$  mM MES pH  $5.5$  in the *trans* side. The reversal potential ( $E_K$ ,  $E_{Na}$ ,  $E_{Cl}$  and  $E_{MES}$ ) for each ion could be calculated ( $E_K = 50.9$  mV,  $E_{Na} = 0$  mV =  $E_{MES}$ , and  $E_{Cl} = -39.5$  mV). In the *trans* side, there is a trace amount of  $Mg^{2+}$  ( $\sim 10$   $\mu$ M), which gives a reversal potential  $E_{Mg}$  of  $92$  mV. Our dye leakage assay showed that the pore was open at  $V_{mem} = 0$  mV transmembrane potential, ruling out significant voltage-dependent gating of the RegIII $\alpha$  channel. On the basis of the ion replacement studies we did for different ions, we estimated the relative permeability of different ions to be:  $P_K = P_{Na} = 1.0$ ;  $P_{Cl} = 0.85$ ;  $P_{MES} = 0.73$  and  $P_{Mg} = 0.66$ . The measured relative permeation rates showed that the pore has very weak cation selectivity, and favours  $K^+/Na^+$  over  $Mg^{2+}$  due to the charge density difference. Under the same assumption, the average conductance ( $\langle g \rangle$ ) of the two basic opening events ( $i1$  and  $i2$ ) could be calculated as the following:

$$\langle g \rangle = \frac{i}{\sum P_{ion}(V_{mem} - E_{ion})}$$

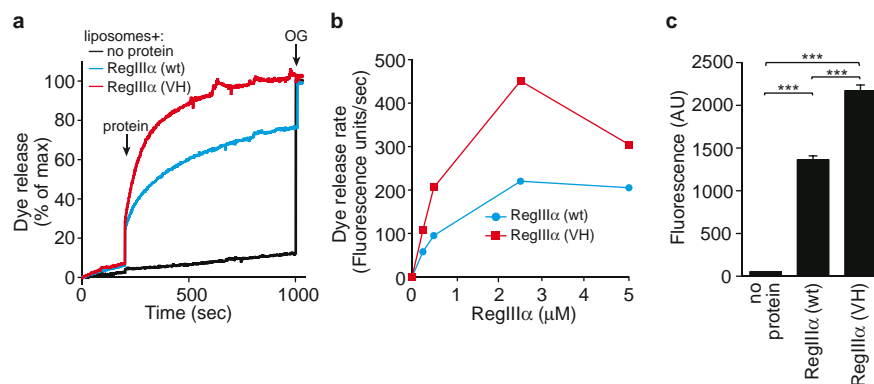
The two calculated conductance levels of  $100$  pS and  $152$  pS were then entered into the Nernst–Planck equation for electrodiffusion and gave rise to an approximate estimate of the pore diameter of  $12$  Å and  $14$  Å, respectively, which is in good agreement with the observed pore size in the reconstructed three-dimensional structure of the pore (Fig. 3b). A more rigorous calculation of the ion flux is possible with a high-resolution picture of the potential profile, but is beyond the scope of this paper.





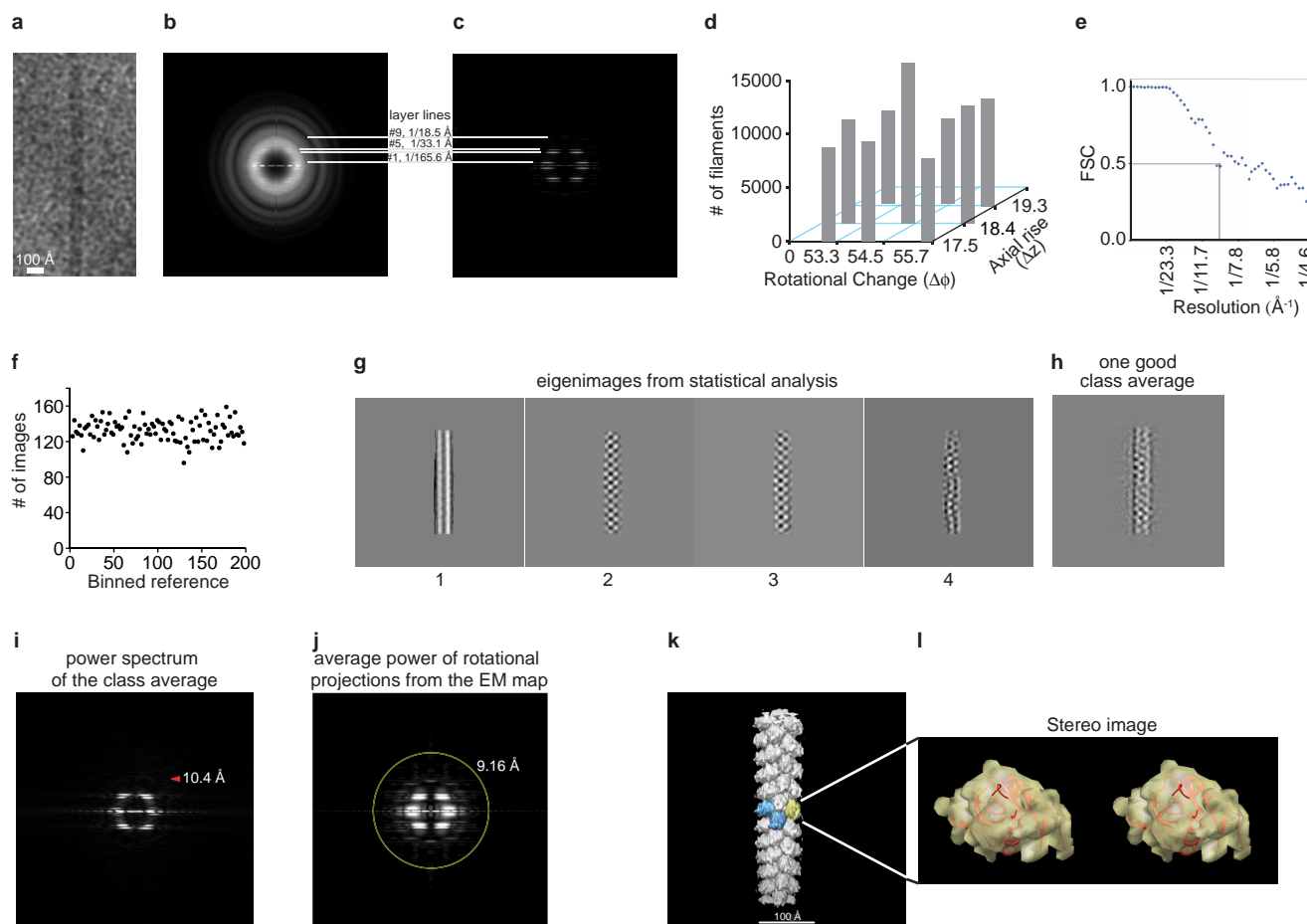
**Extended Data Figure 4 | Analysis of liposome-associated RegIIIα by electron microscopy.** **a**, Negative staining EM controls lacking RegIIIα or liposomes are shown. **b–d**, RegIIIα pore complexes assemble into filaments. **b**, RegIIIα forms filaments in the presence of lipid vesicles. 20 μM RegIIIα was incubated for 2 or 20 min with vesicles composed of PC/PS (85%:15%). Samples were visualized by transmission electron microscopy. Grids were stained with anti-RegIII antibody<sup>1,10</sup> to confirm that the filaments were composed of RegIIIα. Filamentation required membranes, as no filaments were observed in the absence of liposomes. Arrows indicate examples of filaments in each image. **c**, 20 μM RegIIIα carrying a mutation near the C terminus (C-terminal

sequence: FTD (wild-type)→VH (mutant)) was incubated for 20 min with unilamellar vesicles and visualized by cryoEM and negative-staining EM. The results demonstrate that the VH mutant retains the ability to form pores in lipid bilayers but cannot form filaments. A comparison of the wild-type and mutated C terminus is shown below. **d**, Quantification of filament formation by 20 μM pro-RegIIIα, wild-type (wt) and C-terminal mutant (VH) RegIIIα in the presence of vesicles. Results are representative of counts from three different areas. nd, not detected. The results show that pro-RegIIIα, which cannot form pores, also cannot assemble into filaments.



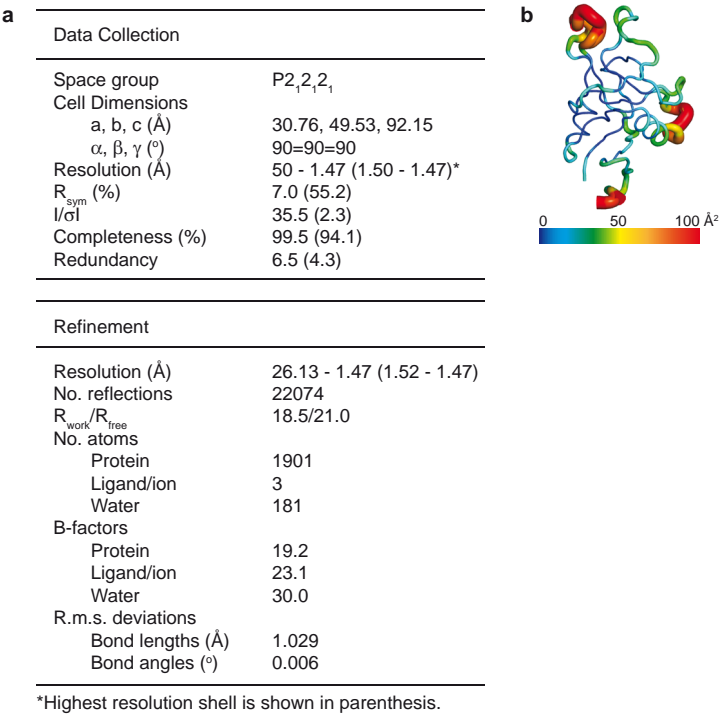
**Extended Data Figure 5 | Filament formation inhibits RegIIIα membrane toxicity.** We examined the functional properties of the RegIIIα VH mutant carrying a mutation near the C terminus (C-terminal sequence: FTD (wild-type)→VH (mutant)), thus truncating the protein near the C terminus. The VH mutant lacks the ability to form filaments but retains the ability to form pores. In accordance with its pore-forming activity, the RegIIIα VH mutant retained membrane toxicity against liposomes and live bacteria. In fact, membrane toxicity was modestly enhanced in the RegIIIα VH mutant, suggesting that trapping of the pore complexes in filaments inhibits their membrane permeabilizing activity. This function contrasts with that of human

α-defensin-6 filaments, which directly trap bacteria in 'nanonets'<sup>20</sup>. **a**, 1.0 μM wild-type (wt) and RegIIIα (VH) mutant was added to 10 μM carboxyfluorescein-loaded liposomes and dye release was monitored. The detergent octylglucoside (OG) was added at the end of the experiment to disrupt remaining liposomes. **b**, Initial rate of liposome dye release (10 μM lipid) as a function of wild-type and mutant RegIIIα concentration. **c**, 5.0 μM wild-type or RegIIIα (VH) mutant was assayed for membrane disruptive activity towards whole bacteria using the SYTOX uptake assay described in Fig. 1. Assays were performed in triplicate. Error bars indicate s.d.; \*\*\* $P < 0.001$ .



**Extended Data Figure 6 | CryoEM reconstruction of the RegIII $\alpha$  filament structure.** **a**, Raw image of a single filament. **b**, **c**, Comparison of the average power spectrum of cryoEM images of individual short helical segments (**b**) and the average power spectrum (**c**) from the projections of the three-dimensional reconstruction at evenly sampled rotation angles around the helical axis. Layer lines 1, 5 and 9 were labelled, and layer line 4 was clearly visible. **d**, Symmetry variability ( $\Delta\phi$  and  $\Delta z$ ) in the cryoEM data set. The reconstruction from the aligned images was imposed with symmetry parameters that vary around the centre pair ( $\Delta\phi = 54.5^\circ$  and  $\Delta z = 18.4$  Å), and the experimental data set was classified into nine bins by projection matching. The populations in these classes were exhibited in a three-dimensional histogram. Even though the central bin is the most populated, the distribution is approximately flat. **e**, Fourier shell correlation (FSC) calculated from the two independent volumes but windowed in different boxes. The strong symmetry in the two volumes led to the FSC  $\sim 0.2$  at the Nyquist frequency. The first fast drop of FSC curve to 0.5 was elected to give an approximate estimate of resolution. **f**, Number of the filament images aligned with each reference projection from the three-dimensional model in the last round of refinement. The projections from the

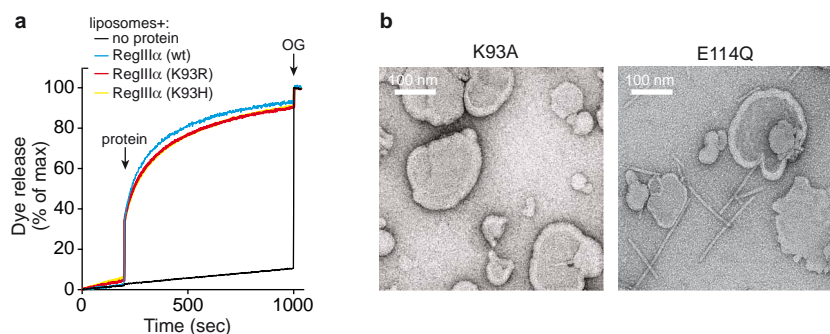
three-dimensional model evenly sampled the orientation space. As expected, the distribution is fairly flat. **g–j**, Statistical analyses of the RegIII $\alpha$  filament structure. **g**, First four eigenimages from the multivariate statistical analysis of the centred filaments in the data set that were padded to 320 pixels in size. The second and third images lack mirror symmetry around the central line, suggesting the parity is odd. The fourth image shows the significant local bending of the filaments, a major limiting factor for us in reaching a better resolution in our reconstruction. **h**, A good class average after the multivariate statistical analysis and hierarchical classification. **i**, Square root of calculated power spectrum of the class average in **h**. The tip of the red arrowhead points at 10.4 Å. **j**, The layer lines in the average power spectrum of the rotational projections from the final reconstruction without symmetry imposition extend isotropically to  $\sim 9.2$  Å (yellow circle), and further along the vertical direction (helical axis). **k**, **l**, Docking of the RegIII $\alpha$  crystal structure into the cryoEM map. **k**, The three-dimensional reconstruction calculated from the images in the central bin, **d**, with a hexameric pore highlighted. **l**, Stereo image showing docking of the RegIII $\alpha$  crystal structure in the cryoEM density map of one subunit out of the reconstruction.



**Extended Data Figure 7 | Crystal structure of bactericidally active RegIIIα.**

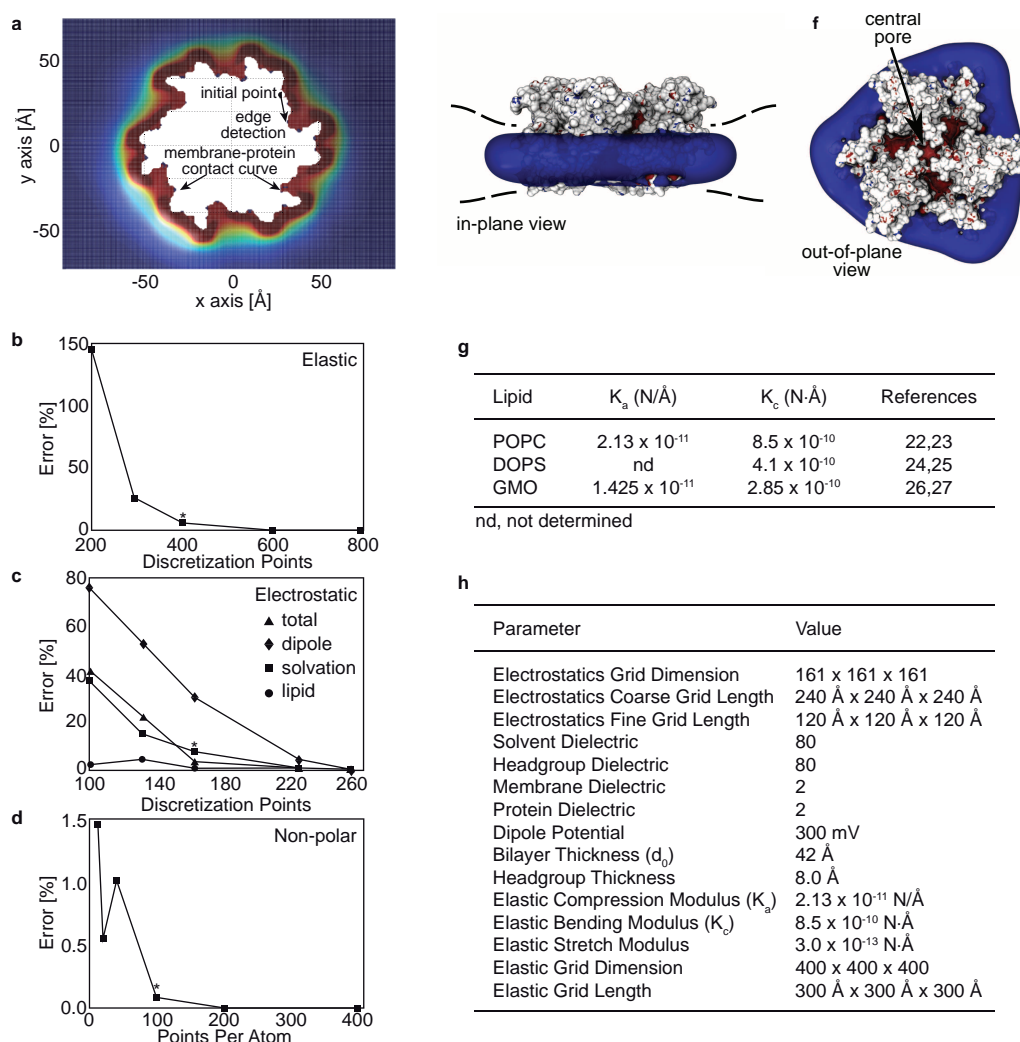
**a**, Table showing data collection and refinement statistics for the active RegIIIα crystal structure. **b**, Crystallographic *B*-factor map of the active RegIIIα structure showing areas of conformational flexibility. Red indicates greater flexibility.





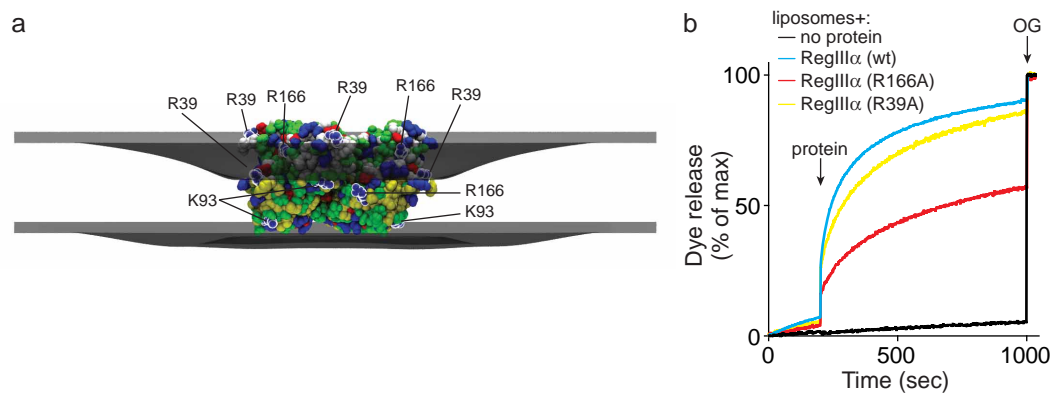
**Extended Data Figure 8 | RegIII $\alpha$  mutagenesis.** **a**, Mutagenesis of Lys 93 (K93) with conservative amino acid substitutions (Arg (R) and His (H)) does not alter membrane toxicity of RegIII $\alpha$ . 5  $\mu$ M of wild-type, Lys93Arg mutant, or Lys93His mutant RegIII $\alpha$  was added to 100  $\mu$ M carboxyfluorescein-loaded liposomes and dye release was monitored. These mutants retain membrane toxicity, in contrast to Lys93Ala (Fig. 3e), suggesting the importance of positive

charges at these sites. **b**, Filamentation of RegIII $\alpha$  mutants (Lys93Ala (K93A) and Glu114Gln (E114Q)) correlates with membrane toxicity. 20  $\mu$ M RegIII $\alpha$  Lys93Ala (left panel) or Glu114Gln (right panel) was incubated for 20 min with unilamellar vesicles and visualized by negative-staining EM. The results demonstrate that the non-toxic Glu114Gln mutant, unlike the toxic Lys93Ala mutant, assembles into filaments.



**Extended Data Figure 9 | Computational modelling of RegIII $\alpha$  insertion into membranes.** **a**, Top-down view of the numeric grid and complex boundary used in the elasticity calculations to represent the upper leaflet. The protein complex occupies the white space in the centre, and the membrane-protein contact curve is the red-white boundary. The membrane is modelled in all non-white regions. The rectangular grid for the elasticity solver is shown here coloured by the membrane bending energy density (red is high bending energy and blue is low bending energy). This calculation corresponds to the membrane bending shown in Fig. 3g. **b–d**, Numeric convergence of the model. **b**, Convergence of the elastostatic energy. In all panels, per cent error was calculated as  $100|(E(n) - E(n_{\max}))|/E(n_{\max})$ , where  $E(n)$  is energy calculated with  $n$  grid points, and  $n_{\max}$  is maximum number of grid points used. The elastic energy converges smoothly as  $n$  increases, and we used  $n = 400$  in both the  $x$  and  $y$  directions for all calculations in the main text, which gives a 5% error. **c**, Convergence of the electrostatic energy. Per cent error of the dipole charge-protein interaction energy (diamonds), protein solvation energy (squares), anionic lipid charge-protein interaction energy (circles) and the total

electrostatic energy (triangles) are shown as a function of the grid discretization. A value of  $n = 161$  was used for the calculations discussed in the main text resulting in a total electrostatic error of 2.5%. **d**, Convergence of the non-polar energy. A discretization of  $n = 100$  points was used for the calculations reported in the main text, and this has a very small error on the order of 0.1%. Values used for calculations in the main text are indicated by an asterisk. **e, f**, Electrostatic potential of the RegIII $\alpha$  pore complex. **e**, In-plane view. The Poisson-Boltzmann equation was solved using APBS after embedding the complex in a low dielectric region mimicking the lipid bilayer<sup>21</sup>. The low dielectric membrane region is deformed corresponding with the lowest energy shape predicted by our physics-based computational model. Positive (blue) isocontours of the electrostatic potential are drawn at  $+5 \text{ kcal mol}^{-1} \text{ e}^{-1}$ . **f**, Out-of-plane view. All details are identical to those in panel **a**. Both positive (blue) and negative (red) isocontours of the electrostatic potential are drawn at  $\pm 5 \text{ kcal mol}^{-1} \text{ e}^{-1}$ . **g**, Table showing bilayer material properties used in the modelling calculations. **h**, Table showing model parameters. References 24–29 are cited in this figure.



**Extended Data Figure 10 | Modelling of RegIII $\alpha$ -membrane interactions.** **a**, RegIII $\alpha$  pore complex model shown from the side. Arg 166 (R166) is located near the water-membrane interface, indicating that it is positioned to interact with the phospholipid head-groups, whereas Arg 39 is predicted to be exposed to aqueous solvent. Membrane boundaries predicted from the

computational calculations are indicated. **b**, 5  $\mu$ M of wild-type, Arg166Ala mutant, or Arg39Ala mutant RegIII $\alpha$  was added to 100  $\mu$ M carboxyfluorescein-loaded liposomes and dye release was monitored. The experimental results are consistent with the position of these residues relative to the membrane interface in the model.

20. Chu, H. *et al.* Human  $\alpha$ -defensin 6 promotes mucosal innate immunity through self-assembled peptide nanonets. *Science* **337**, 477–481 (2012).
21. Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 351–371 (1973).
22. Henriksen, J. *et al.* Universal behavior of membranes with sterols. *Biophys. J.* **90**, 1639–1649 (2006).
23. Kučerka, N., Tristram-Nagle, S. & Nagle, J. F. Closer look at structure of fully hydrated fluid phase DPPC bilayers. *Biophys. J.* **90**, L83–L85 (2006).
24. Petrache, H. I. *et al.* Structure and fluctuations of charged phosphatidylserine bilayers in the absence of salt. *Biophys. J.* **86**, 1574–1586 (2004).
25. Fuller, N., Benatti, C. R. & Rand, R. P. Curvature and bending constants for phosphatidylserine-containing membranes. *Biophys. J.* **85**, 1667–1674 (2003).
26. Nielsen, C., Goulian, M. & Andersen, O. S. Energetics of inclusion-induced bilayer deformations. *Biophys. J.* **74**, 1966–1983 (1998).
27. White, S. H. Formation of 'solvent-free' black lipid bilayer membranes from glyceryl monooleate dispersed in squalene. *Biophys. J.* **23**, 337–347 (1978).



# Structural basis for recognition of synaptic vesicle protein 2C by botulinum neurotoxin A

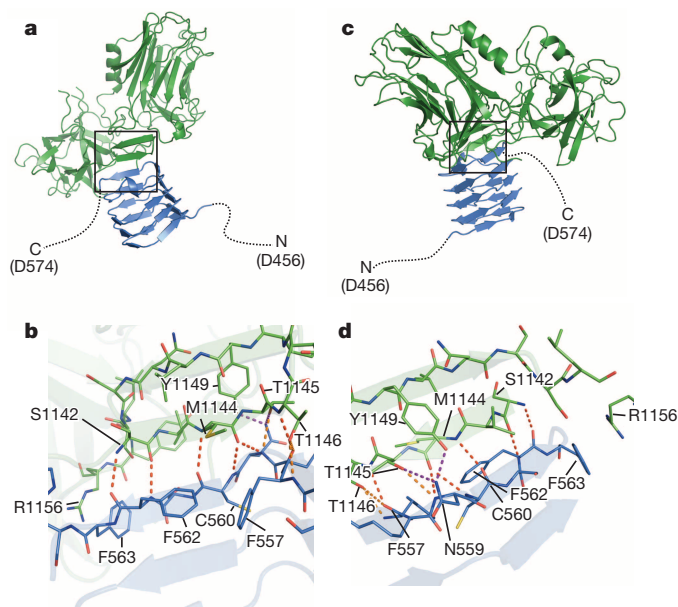
Roger M. Benoit<sup>1</sup>, Daniel Frey<sup>1\*</sup>, Manuel Hilbert<sup>1\*</sup>, Josta T. Kevenaar<sup>2\*</sup>, Mara M. Wieser<sup>1</sup>, Christian U. Stirnimann<sup>3</sup>, David McMillan<sup>4</sup>, Tom Ceska<sup>4</sup>, Florence Lebon<sup>5</sup>, Rolf Jaussi<sup>1</sup>, Michel O. Steinmetz<sup>1</sup>, Gebhard F. X. Schertler<sup>1,6</sup>, Casper C. Hoogenraad<sup>2</sup>, Guido Capitani<sup>1</sup> & Richard A. Kammerer<sup>1</sup>

Botulinum neurotoxin A (BoNT/A) belongs to the most dangerous class of bioweapons<sup>1</sup>. Despite this, BoNT/A is used to treat a wide range of common medical conditions such as migraines and a variety of ocular motility and movement disorders<sup>2</sup>. BoNT/A is probably best known for its use as an antiwrinkle agent in cosmetic applications (including Botox and Dysport)<sup>3</sup>. BoNT/A application causes long-lasting flaccid paralysis of muscles through inhibiting the release of the neurotransmitter acetylcholine by cleaving synaptosomal-associated protein 25 (SNAP-25) within presynaptic nerve terminals<sup>4</sup>. Two types of BoNT/A receptor have been identified, both of which are required for BoNT/A toxicity and are therefore likely to cooperate with each other<sup>5</sup>: gangliosides and members of the synaptic vesicle glycoprotein 2 (SV2) family, which are putative transporter proteins that are predicted to have 12 transmembrane domains, associate with the receptor-binding domain of the toxin<sup>5</sup>. Recently, fibroblast growth factor receptor 3 (FGFR3) has also been reported to be a potential BoNT/A receptor<sup>6</sup>. In SV2 proteins, the BoNT/A-binding site has been mapped to the luminal domain<sup>7</sup>, but the molecular details of the interaction between BoNT/A and SV2 are unknown. Here we determined the high-resolution crystal structure of the BoNT/A receptor-binding domain (BoNT/A-RBD) in complex with the SV2C luminal domain (SV2C-LD). SV2C-LD consists of a right-handed, quadrilateral  $\beta$ -helix that associates with BoNT/A-RBD mainly through backbone-to-backbone interactions at open  $\beta$ -strand edges, in a manner that resembles the inter-strand interactions in amyloid structures. Competition experiments identified a peptide that inhibits the formation of the complex. Our findings provide a strong platform for the development of novel antitoxin agents and for the rational design of BoNT/A variants with improved therapeutic properties.

The overall structure of BoNT/A-RBD in complex with SV2C-LD is shown in Fig. 1, and the data collection and refinement statistics are listed in Extended Data Table 1. The convex interface formed by the amino- and carboxy-terminal BoNT/A-RB subdomains interacts with SV2C-LD. SV2C-LD forms a right-handed, quadrilateral  $\beta$ -helix, a fold that is characteristic of pentapeptide-repeat proteins<sup>8</sup> (Extended Data Fig. 1). Based on a search of the Dali server<sup>9</sup>, the proteins that are most structurally similar to SV2C-LD in the Protein Data Bank (PDB)<sup>10</sup> are MfpA, a pentapeptide-repeat protein from *Mycobacterium tuberculosis*<sup>11</sup>, and AlbG, an McbG-like protein from *Xanthomonas albilineans*<sup>12</sup>.

It has been noted that the molecular architecture of  $\beta$ -helices is similar to the structure of amyloid fibrils<sup>13</sup>. Like in amyloid fibrils, the regular  $\beta$ -sheet edges of  $\beta$ -helices are in a conformation that provides the possibility of interacting with other  $\beta$ -strands and aggregating into higher-order structures. To avoid the formation of such aggregates, the ends of  $\beta$ -helices are typically capped by other secondary-structure

elements<sup>14,15</sup>. No additional secondary-structure elements are seen on the N-terminal side of SV2C-LD, but the first N-terminal turn of the SV2C-LD  $\beta$ -helix contains more charged amino acids in the inward facing positions than do the other coils. Such strategically placed charges are known to prevent aggregation of single-sheet proteins and  $\beta$ -propellers<sup>15</sup>. At its C terminus, SV2C-LD contains a short  $3_{10}$ -helix (Extended Data Fig. 1), but this helix precedes the two  $\beta$ -strands that interact with BoNT/A-RBD. The interaction between the open  $\beta$ -strand of SV2C-LD and the  $\beta$ -strand edge of BoNT/A-RBD is dominated by backbone-backbone hydrogen bonds, resembling the inter-strand interactions in amyloid structures<sup>16,17</sup>. Our findings therefore indicate that *Clostridium botulinum* has exploited a weakness in the structure of SV2 proteins to transport BoNT/A into host cells.



**Figure 1 | The BoNT/A-RBD-SV2C-LD complex.** **a**, Cartoon representation of the structure of BoNT/A-RBD (green) in complex with SV2C-LD (blue). Dotted lines indicate flexible regions of SV2C-LD that were not visible in the structure. The terminal residues of the SV2C-LD expression construct are indicated. **b**, Close-up view of the interaction site (boxed area from **a**). Backbone-backbone and backbone-side chain hydrogen bonds are indicated by dashed orange lines. Side chain-side chain interactions are shown as dashed purple lines. **c**, The complex structure from a different view. **d**, Close-up of the interaction site (boxed region from **c**). **b**, **d**, Carbon is shown in light blue for SV2C-LD and green for BoNT/A-RBD. In both structures, dark blue denotes nitrogen, red denotes oxygen and yellow denotes sulphur.

<sup>1</sup>Laboratory of Biomolecular Research, Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland. <sup>2</sup>Cell Biology, Faculty of Science, Utrecht University, 3584 CH Utrecht, The Netherlands. <sup>3</sup>Swiss Light Source, Paul Scherrer Institut, CH-5232 Villigen PSI, Switzerland. <sup>4</sup>UCB Celltech, UCB Pharma, UCB NewMedicines, Slough SL1 4EN, UK. <sup>5</sup>UCB Pharma, UCB NewMedicines, B-1420 Braine-l'Alleud, Belgium. <sup>6</sup>Department of Biology, ETH Zurich, CH-8093 Zurich, Switzerland.

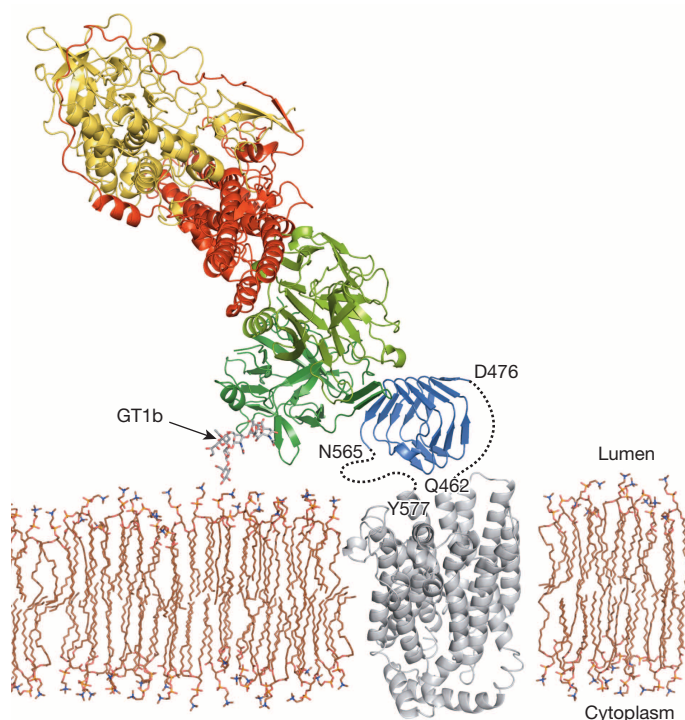
\*These authors contributed equally to this work.

BoNT/A has been shown to bind to neurons not only through SV2 proteins<sup>7</sup>, or potentially FGFR3 (ref. 6), but also in conjunction with gangliosides as co-receptors<sup>18</sup>. A superimposition of our BoNT/A-RBD–SV2C-LD complex crystal structure with the previously reported full-length BoNT/A structure<sup>19</sup> and the BoNT/A-RBD–GT1b structure<sup>18</sup> shows that the binding site for gangliosides is located in the C-terminal BoNT/A-RB subdomain almost opposite to the SV2C-binding site (Fig. 2). The two binding regions are separated by a distance of approximately 40 Å, illustrating that both receptors can simultaneously interact with the toxin. This finding might have important functional implications for the translocation of BoNT/A across the synaptic vesicle membrane and for subsequent cell intoxication, which are mechanisms of major importance that are still poorly understood. Recent studies using BoNT/B and BoNT/E have demonstrated that the low pH in the endosomal lumen and binding to the ganglioside GT1b are both required for the toxins to transform into oligomeric, hydrophobic membrane-associated channels<sup>20,21</sup>. Although the mechanism of channel formation has not yet been studied at such detail for BoNT/A, additional binding to SV2 proteins might be a prerequisite for toxin translocation. Alternatively, binding to SV2 proteins might prevent the translocation of BoNT/A by hindering the formation of oligomeric transmembrane channels. Such inhibition might be relieved by acidification in the endosomal lumen and result both in the dissociation of the SV2–BoNT/A complex and the known inhibition of BoNT/A translocation by the RBD at neutral pH<sup>22</sup>. One promising candidate for a pH-sensing residue in SV2C-LD is H564, which is present at the interface of the complex. Consistent with this hypothesis, we observed an approximately five-fold decrease in the dissociation constant ( $K_d$ ) of the complex between

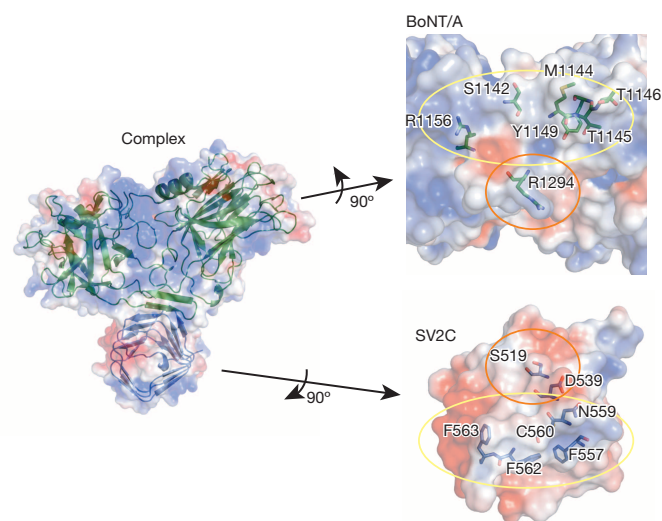
pH 7.5 and pH 5 in fluorescence anisotropy experiments (Extended Data Fig. 2).

Comparison of BoNT/A-RBD structures in the free state and SV2C-LD-bound state reveals no significant conformational changes in the SV2C-LD-interacting elements of BoNT/A-RBD on complex formation.  $\alpha$  superimposition of the BoNT/A-RBD<sup>23</sup> onto the BoNT/A-RBD–SV2C-LD complex results in a root mean squared deviation (r.m.s.d.) of 0.54 Å over 412 residues. At the interface, 15 SV2C-LD residues and 19 BoNT/A-RBD residues are partially or fully buried, resulting in a contact area of approximately 596 Å<sup>2</sup>. Several of these buried residues are engaged in backbone–backbone hydrogen bonds between interacting  $\beta$ -strands (Extended Data Table 2). One prominent feature seen in the structure is the accumulation of positively charged surface residues at the BoNT/A-RBD interface, whereas the binding surface of SV2C-LD is slightly negatively charged (Fig. 3). The residues are not engaged in inter-domain salt bridges; therefore, relatively few specific side-chain interactions are observed in the complex structure. Among these, the interaction between R1156 of BoNT/A-RBD and F563 of SV2C-LD is of particular interest, as it is a cation– $\pi$ -stacking interaction. The hydrogen bond between SV2C-LD N559 and BoNT/A-RBD Y1149 extends the interaction to the more distant  $\beta$ -strand of the toxin. Hydrogen bonds between SV2C-LD N559 and BoNT/A-RBD T1145, as well as between the backbone oxygen and nitrogen of SV2C-LD F557 and BoNT/A-RBD T1146, stabilize the complex downstream of the C-terminal boundary of the directly interacting toxin  $\beta$ -strand. R1294 is a residue that prominently contributes to the positive charge of the BoNT/A-RBD surface. Together, these residues seem to be important for shape and charge complementarity and consequently for binding specificity.

The importance of these prominent side-chain interactions was tested by site-directed mutagenesis (Extended Data Fig. 3). Mutation of the two BoNT/A-RBD threonines T1145 and T1146 completely abrogated binding to SV2C-LD in pull-down assays, and mutation of R1294 or R1156 in BoNT/A-RBD significantly reduced this binding (Extended Data Fig. 3). Likewise, substitution of the phenylalanine at position 563 of SV2C-LD with alanine strongly reduced the binding to

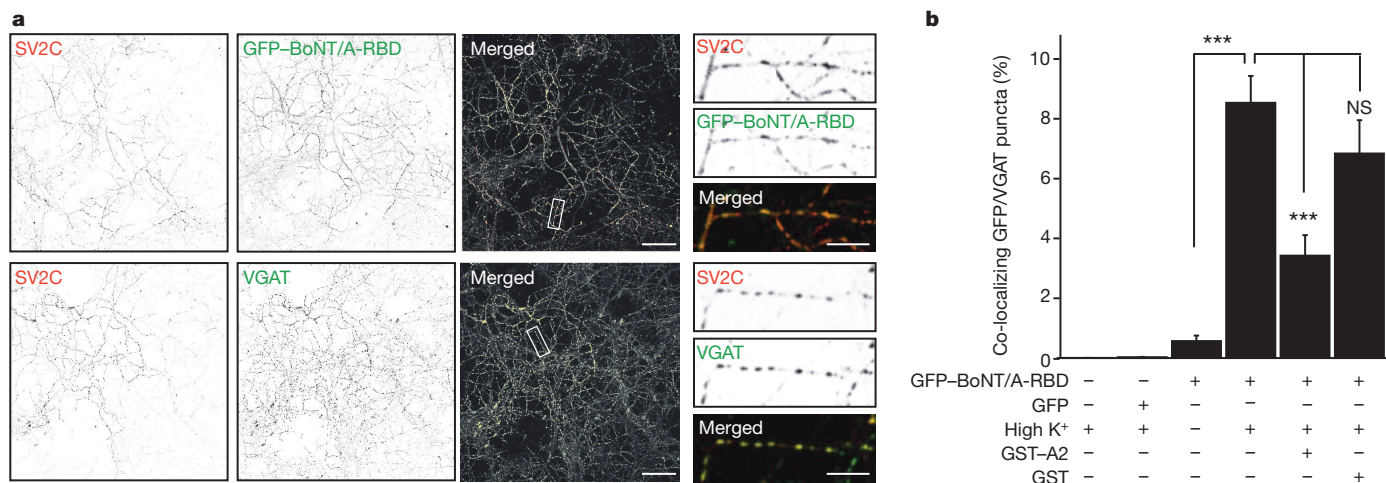


**Figure 2 | Proposed model for simultaneous ganglioside and SV2C binding by BoNT/A.** Superimposition of full-length BoNT/A<sup>19</sup>, the BoNT/A-RBD–GT1b ganglioside complex<sup>18</sup> and the BoNT/A-RBD–SV2C-LD complex. The ganglioside (white and atom colours; labelled in the figure) and SV2C-LD (blue) bind to opposite sides of the C-terminal BoNT/A-RB subdomain (dark green). The N-terminal BoNT/A-RB subdomain (pale green), the translocation domain (including the belt (red)) and the protease domain (yellow) point away from the binding site. The SV2C transmembrane portion (light grey) was modelled. The position of the plasma membrane is proposed. Linker regions are shown as black dotted lines. Linker boundaries are indicated.



**Figure 3 | Open book representation of the BoNT/A-RBD–SV2C-LD interaction site.** The components of the complex are shown separately and coloured according to their electrostatic potential ( $\pm 3 kT/e$ , where  $k$  is the Boltzmann constant,  $T$  is temperature and  $e$  is the elementary charge): blue denotes basic residues, red denotes acidic residues and white denotes hydrophobic residues. The yellow ellipse indicates the approximate location of the hydrogen-bonding residues. The orange ellipse indicates the positively charged R1294 in BoNT/A-RBD and the mostly negatively charged surface region in SV2C-LD. R1294 is not defined in the electron density of the complex structure and hence does not participate in hydrogen bonds or salt bridges.





**Figure 4 | The BoNT/A-A2 peptide inhibits the internalization of BoNT/A-RBD by striatal neurons (at 18 days *in vitro*).** **a**, Co-localization of endogenous SV2C with GFP-BoNT/A-RBD (top) or endogenous vesicular inhibitory amino acid transporter (VGAT) (bottom) in cultured striatal neurons. Scale bars, 50  $\mu$ m. Enlargements of the regions indicated by the white rectangles are shown in the panels on the right. Scale bar, 10  $\mu$ m. **b**, Quantification of GFP-BoNT/A-RBD uptake by cultured striatal neurons. The number of co-localizing GFP-BoNT/A-RBD and VGAT puncta was

normalized to the total number of VGAT puncta. No GFP-BoNT/A-RBD, 0.01%  $\pm$  0.00%; GFP, 0.03%  $\pm$  0.01%; no high K<sup>+</sup> buffer, 0.57%  $\pm$  0.18%; no glutathione S-transferase (GST)-BoNT/A-A2 (GST-A2), 8.53%  $\pm$  0.09%; GST-BoNT/A-A2 (GST-A2), 3.42%  $\pm$  0.69%; GST, 6.84%  $\pm$  1.12%.  $n = 20$  images per group from two cultures, Mann-Whitney  $U$  test; \*\*\* $P < 0.001$ ; NS, not significant.  $N = 2$  independent experiments. All tests were performed two-sided. Data are presented as mean  $\pm$  s.e.m.

BoNT/A-RBD, whereas mutation of the SV2C-LD residue N559 had no significant effect on interaction with the toxin. To further characterize the complex interaction, we performed fluorescence anisotropy experiments. Analysis of the wild-type proteins yielded a  $K_d$  value of  $0.26 \pm 0.2 \mu$ M.  $K_d$  values obtained for the mutant proteins were consistent with the pull-down results (Extended Data Fig. 4 and Extended Data Table 3).

It has been reported that SV2 peptides harbouring the toxin-binding site prevent BoNT/A from binding to neurons<sup>7</sup>. We therefore analysed the potency of short BoNT/A-RBD- and SV2C-LD-derived peptides at inhibiting complex formation. Based on the interactions seen in the crystal structure, the peptide sequences were designed to correspond to the two SV2C-LD  $\beta$ -strands (peptide SV2C-A3), the BoNT/A-RBD  $\beta$ -strand (peptide BoNT/A-A5) and the BoNT/A-RBD  $\beta$ -sheet (peptide BoNT/A-A2) that interact in the complex (Extended Data Fig. 3g). We observed full inhibition of complex formation in the presence of the peptide corresponding to the BoNT/A-RBD  $\beta$ -sheet and partial inhibition in the presence of the peptide corresponding to the two SV2C-LD  $\beta$ -strands, whereas no reduction in binding was seen in the presence of the peptide corresponding to the single BoNT/A-RBD  $\beta$ -strand. These results indicate that the three-dimensional conformation of SV2C-LD is crucial for the interaction. Consistent with this conclusion, the shortest SV2C-LD fragment reported to bind to BoNT/A spans residues 529–566 of the SV2C sequence<sup>7</sup>. The fragment extends over almost two turns of the  $\beta$ -helix and might therefore form a moderately stable three-dimensional structure. We next tested the BoNT/A-A2  $\beta$ -sheet peptide for inhibition of BoNT/A-RBD binding to and internalization by cultured striatal neurons (at 18 days *in vitro*) (Fig. 4 and Extended Data Fig. 5), which endogenously express SV2C, or HEK-293 cells that were transfected with Flag-tagged SV2C protein (Extended Data Fig. 6). In the absence of the toxin-derived peptide, BoNT/A-RBD co-localized with SV2C. Pre-incubation of the cells with 5  $\mu$ M glutathione S-transferase (GST)-tagged BoNT/A-RBD peptide resulted in a significant decrease in toxin binding and internalization, whereas no inhibition of the complex interaction was observed in the presence of a GST-only control.

The findings of our study have important implications for the development of novel antitoxin agents, as well as for the medical and cosmetic applications of BoNT/A. The toxin has been described as one of the six most dangerous bioweapons<sup>1</sup>. Currently, the only available antidotes

for BoNT/A have severe side effects. Proposed alternative approaches include peptide inhibitors, and all three domains of the BoNTs have been considered as potential drug targets<sup>24</sup>. The detailed structural information reported on the SV2C-LD-BoNT/A-RBD interaction should therefore be of particular interest for the targeted development of BoNT/A-specific antibodies or high-affinity peptide inhibitors directed against the SV2-binding interface.

Depending on the type and area of medical or cosmetic application, the recommended dose of BoNT/A varies markedly<sup>25</sup>, carrying a substantial risk of accidental BoNT/A overdose and poisoning. The crystal structure of the BoNT/A-RBD in complex with SV2C-LD provides a strong structural basis for the rational design of BoNT/A variants with an attenuated capacity to bind to SV2 proteins. Such variants are promising candidate proteins for broadening the very narrow therapeutic window, possibly resulting in much safer applications of the toxin.

## METHODS SUMMARY

All of the methods are described in detail in the Methods. The proteins were cloned into pET-based vectors and expressed in bacteria using auto-induction. Protein purification was performed using immobilized metal affinity chromatography (IMAC) or GST sepharose affinity chromatography followed by size-exclusion chromatography. The purified BoNT/A and SV2C domains were combined in a 1:1 ratio. The complex was purified in an additional size-exclusion chromatography step, concentrated to 8 mg ml<sup>-1</sup> and crystallized by vapour diffusion at 20 °C. The reservoir solution consisted of 100 mM HEPES, pH 7.5, 6% PEG 8000, 8% glycerol and 100 mM NaCl. For cryoprotection, the reservoir solution was supplemented with an additional 10% glycerol. Diffraction data were collected at the Swiss Light Source (SLS). The crystals belong to space group C 1 2 1 and are pseudomerohedrally twinned with twin law  $-h, -k, l$ . The structure was solved by molecular replacement using PDB entry 3FUO as a search model. Pull-down assays were performed using IMAC with 6 $\times$ His-tagged BoNT/A-RBD and untagged SV2C-LD.  $K_d$  determination was performed by fluorescence anisotropy. For functional assays, primary striatal cultures were prepared from embryonic day 18 rat brains, as described for hippocampal neurons<sup>26</sup>. At 18 days *in vitro*, cultured striatal neurons were stimulated using a high K<sup>+</sup> buffer (70 mM KCl, 51.5 mM NaCl, 25 mM HEPES, 30 mM glucose, 2 mM CaCl<sub>2</sub> and 2 mM MgCl<sub>2</sub>; pH 7.4) with or without the GST-BoNT/A-A2 peptide or a GST control, or neurons were treated with a control buffer (2.5 mM KCl, 119 mM NaCl, 25 mM HEPES, 30 mM glucose, 2 mM CaCl<sub>2</sub> and 2 mM MgCl<sub>2</sub>; pH 7.4), for 15 min at 37 °C. The neurons were incubated with GFP-BoNT/A-RBD or GFP only for 10 min at 37 °C and then washed with high K<sup>+</sup> or control buffer and then fixed. HEK293T cells were cultured in DMEM/Ham's F-10 (50/50) medium supplemented with 10% FCS

and 1% penicillin/streptomycin at 37 °C in 5% CO<sub>2</sub> and were transfected with the plasmid pGW2-mRFP (monomeric red fluorescent protein) and SV2C-Flag, or pGW2-mRFP only. After 24 h, the cells were incubated with fresh culture medium containing GST-BoNT/A-A2 or a GST control for 15 min at 37 °C. The cells were then incubated with BoNT/A-RBD for 10 min at 37 °C, washed with fresh culture medium and fixed. Immunohistochemistry, confocal imaging, data representation and statistical analyses were performed as previously described<sup>27</sup>. Image analysis was performed using ImageJ software.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 11 January; accepted 3 October 2013.**

**Published online 17 November 2013.**

- Arnon, S. S. *et al.* Botulinum toxin as a biological weapon: medical and public health management. *J. Am. Med. Assoc.* **285**, 1059–1070 (2001).
- Carruthers, J. & Carruthers, A. Botox: beyond wrinkles. *Clin. Dermatol.* **22**, 89–93 (2004).
- Markey, A. C. Botulinum A exotoxin in cosmetic dermatology. *Clin. Exp. Dermatol.* **25**, 173–175 (2000).
- Montecucco, C. & Schiavo, G. Mechanism of action of tetanus and botulinum neurotoxins. *Mol. Microbiol.* **13**, 1–8 (1994).
- Montal, M. Botulinum neurotoxin: a marvel of protein design. *Annu. Rev. Biochem.* **79**, 591–617 (2010).
- Jacky, B. P. *et al.* Identification of fibroblast growth factor receptor 3 (FGFR3) as a protein receptor for botulinum neurotoxin serotype A (BoNT/A). *PLoS Pathogens* **9**, e1003369 (2013).
- Dong, M. *et al.* SV2 is the protein receptor for botulinum neurotoxin A. *Science* **312**, 592–596 (2006).
- Vetting, M. W. *et al.* Pentapeptide repeat proteins. *Biochemistry* **45**, 1–10 (2006).
- Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
- Bernstein, F. C. *et al.* The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542 (1977).
- Hegde, S. S. *et al.* A fluoroquinolone resistance protein from *Mycobacterium tuberculosis* that mimics DNA. *Science* **308**, 1480–1483 (2005).
- Vetting, M. W., Hegde, S. S., Zhang, Y. & Blanchard, J. S. Pentapeptide-repeat proteins that act as topoisomerase poison resistance factors have a common dimer interface. *Acta Crystallogr. F* **67**, 296–302 (2011).
- Krishnan, R. & Lindquist, S. L. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature* **435**, 765–772 (2005).
- Bryan, A. W. Jr, Starner-Kreinbrink, J. L., Hosur, R., Clark, P. L. & Berger, B. Structure-based prediction reveals capping motifs that inhibit  $\beta$ -helix aggregation. *Proc. Natl Acad. Sci. USA* **108**, 11099–11104 (2011).
- Richardson, J. S. & Richardson, D. C. Natural  $\beta$ -sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl Acad. Sci. USA* **99**, 2754–2759 (2002).
- Ritter, C. *et al.* Correlation of structural elements and infectivity of the HET-s prion. *Nature* **435**, 844–848 (2005).
- Wasmer, C. *et al.* Amyloid fibrils of the HET-s(218–289) prion form a  $\beta$  solenoid with a triangular hydrophobic core. *Science* **319**, 1523–1526 (2008).
- Stenmark, P., Dupuy, J., Imamura, A., Kiso, M. & Stevens, R. C. Crystal structure of botulinum neurotoxin type A in complex with the cell surface co-receptor GT1b—insight into the toxin–neuron interaction. *PLoS Pathogens* **4**, e1000129 (2008).
- Lacy, D. B., Tepp, W., Cohen, A. C., DasGupta, B. R. & Stevens, R. C. Crystal structure of botulinum neurotoxin type A and implications for toxicity. *Nature Struct. Biol.* **5**, 898–902 (1998).
- Sun, S., Tepp, W. H., Johnson, E. A. & Chapman, E. R. Botulinum neurotoxins B and E translocate at different rates and exhibit divergent responses to GT1b and low pH. *Biochemistry* **51**, 5655–5662 (2012).
- Sun, S. *et al.* Receptor binding enables botulinum neurotoxin B to sense low pH for translocation channel assembly. *Cell Host Microbe* **10**, 237–247 (2011).
- Fischer, A., Mushrush, D. J., Lacy, D. B. & Montal, M. Botulinum neurotoxin devoid of receptor binding domain translocates active protease. *PLoS Pathogens* **4**, e1000245 (2008).
- Fu, Z., Chen, C., Barbieri, J. T., Kim, J. J. & Baldwin, M. R. Glycosylated SV2 and gangliosides as dual receptors for botulinum neurotoxin serotype F. *Biochemistry* **48**, 5631–5641 (2009).
- Cai, S. & Singh, B. R. Strategies to design inhibitors of *Clostridium botulinum* neurotoxins. *Infect. Disord. Drug Targets* **7**, 47–57 (2007).
- Anderson, E. R. Jr. Proper dose, preparation, and storage of botulinum neurotoxin serotype A. *Am. J. Health Syst. Pharm.* **61**, S24–S29 (2004).
- Kaech, S. & Banker, G. Culturing hippocampal neurons. *Nature Protocols* **1**, 2406–2415 (2006).
- Spangler, S. A. *et al.* Liprin- $\alpha$ 2 promotes the presynaptic recruitment and turnover of RIM1/CASK to facilitate synaptic transmission. *J. Cell Biol.* **201**, 915–928 (2013).

**Acknowledgements** We thank A. Blanc for performing the mass spectrometry analysis of the recombinant proteins and L. Knecht for help with protein production. This work was supported by UCB Pharma, UCB NewMedicines. This work was also supported by the Swiss National Science Foundation (grant 310030B\_138659), The Netherlands Organization for Scientific Research (NWO-ALW-VICI) and The Netherlands Organization for Health Research and Development (ZonMW-TOP).

**Author Contributions** R.M.B., M.H., D.M., T.C., F.L., M.O.S., G.F.X.S., C.C.H. and R.A.K. designed the research. R.M.B., D.F., M.H., J.T.K., M.M.W., C.U.S., R.J. and G.C. carried out the research. R.M.B., M.H., J.T.K., C.C.H., G.C. and R.A.K. analysed the data. R.M.B. and R.A.K. wrote the paper with input from the other authors.

**Author Information** Atomic coordinates for BoNT/A-RBD–SV2C-LD have been deposited in the Protein Data Bank database under accession number 4JRA. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.A.K. ([richard.kammerer@psi.ch](mailto:richard.kammerer@psi.ch)).



## METHODS

**Cloning and protein production.** The cDNAs encoding human SV2C (Swiss-Prot Q496J9) amino acids 456–574 and BoNT/A (Swiss-Prot P10845) amino acids 871–1296 were amplified by PCR using DNA templates that were codon-optimized for expression in *Escherichia coli* (GENEWIZ). The primer sequences used were: BoNT/A-RBD, 5'-GTGCCGCGCGGATCCAAAACATCATCAACACGTCCATTCTCAACCTC-3' and 5'-TTGCTAAGTGAGCTCTCATACAGCGGACGTTACACCCCAACC-3'; and SV2C-LD, 5'-GTGCCGCGCGGATCCGACGTAATCAAGCCGCTCCAGTC-3' and 5'-TTGCTAAGTGAGCTCTCATATAGTCAAACGTGATTTGGCAACCCG-3'. The BoNT/A-A2 ( $\beta$ -sheet) peptide insert encoding residues 1140–1153 was generated by using synthetic primers (5'-CCACGAACATATATCTGAATTCAGCTGATAATTGACAGAGCTCATTAGCAAGATATAATACAAATCCGCGGAGC-3' and 5'-CAGATATATGTTCTGTTGTCATTACACTACCGCGACCGGATCCGGTCCCTGGAACAGAACTCCAGATCCGATTTTG-3'). Cloning into in-house assembled, pET-based vectors containing an N-terminal 6 $\times$ His, 6 $\times$ His-GFP or GST tag was performed by co-transformation cloning<sup>28</sup>. Mutagenesis was performed as described previously<sup>29</sup>. All insert sequences were verified by DNA sequencing (by GATC).

The proteins were expressed separately in *E. coli* strain NiCo21(DE3) (New England Biolabs). Precultures (100 ml Luria Bertani (LB) medium supplemented with 50  $\mu$ g ml<sup>-1</sup> kanamycin) were grown overnight at 37 °C in a shaking incubator. ZYM-5052 medium for auto-induction<sup>30</sup> (4 l supplemented with 50  $\mu$ g ml<sup>-1</sup> kanamycin) was inoculated 1:40 with the overnight culture, and the bacteria were grown for 7 h at 37 °C. Subsequently, the temperature was lowered to 20 °C, and the incubation was continued overnight. The GST control was expressed in LB medium supplemented with 100  $\mu$ g ml<sup>-1</sup> ampicillin, using induction with 1 mM isopropylthiogalactoside (IPTG).

The cells were harvested by centrifugation. The cell pellets were resuspended in 100 ml lysis buffer (50 mM Tris, pH 7.5, 500 mM NaCl, 10 mM imidazole, 10 mM  $\beta$ -mercaptoethanol and 1 cOmplete EDTA-free protease inhibitor cocktail tablet (Roche Diagnostics)). The cells were lysed on ice by ultrasonication. Lysate clearing was performed for 1 h at 25,000g, and the resultant supernatant was filtered (0.45  $\mu$ m filter). The proteins were subsequently purified by IMAC (on a 5 ml HisTrap FF Crude column, GE Healthcare) and gel filtration (HiLoad 16/60 Superdex 200, GE Healthcare). The GST control was purified by GST sepharose affinity chromatography (GE Healthcare) followed by size-exclusion chromatography. SV2C-LD and BoNT/A-RBD were combined in a 1:1 ratio, co-purified by size-exclusion chromatography and concentrated to 8 mg ml<sup>-1</sup> for crystallization.

**Crystallization and structure elucidation.** The BoNT/A-RBD–SV2C-LD complex was crystallized in 100 mM HEPES, pH 7.5, 6% PEG 8000, 8% glycerol and 100 mM NaCl using a grid screen around the conditions described for the crystallization of BoNT/A-RBD alone<sup>23</sup>. For cryoprotection, the reservoir solution was supplemented with an additional 10% glycerol. A complete 2.3 Å resolution data set was collected from a single crystal at 100 K at the X06DA beamline of the Swiss Light Source using a wavelength of 1.000 Å. The high-resolution limit was chosen according to the guidelines recently suggested by Karplus and Diederichs<sup>31</sup>. Data processing was carried out with XDS<sup>32</sup>. The crystals belong to space group C 1 2 1 and are pseudomerohedrally twinned with twin law  $-h, -k, l$ . The structure of the BoNT/A-RBD–SV2C-LD complex was solved by molecular replacement using residues 875–1295 of BoNT/A-RBD from PDB entry 3FUO as a search model<sup>23</sup>. Two BoNT/A-RBD molecules could be positioned in the C 1 2 1 asymmetric unit by using Phaser<sup>33</sup>. The initial atomic model was refined with phenix.refine<sup>34</sup>, taking into account the twin law. From the first refinement cycles, additional density corresponding to SV2C-LD became visible. This allowed manual building of two copies of the domain, each one bound to a copy of BoNT/A-RBD in a 1:1 complex. Coot<sup>35</sup> was used for model building. The final refined structure exhibits good geometry and stereochemistry, as validated with MolProbity<sup>36</sup> (residues in favoured regions 90.17% and outliers 0.81%). Structure figures were prepared using PyMOL<sup>37</sup>. Electrostatic potentials were calculated using APBS<sup>38</sup>. Secondary-structure matching superimpositions<sup>39</sup> were performed using Coot<sup>35</sup>. PROMOTIF<sup>40</sup> was used for structure analysis. A homology model of the transmembrane part of SV2C was generated using Phyre<sup>41</sup>. A PDB file of a lipid bilayer slice by Heller *et al.*<sup>42</sup> was used for the generation of Fig. 2. Hydrogen bonds at the complex interface were analysed using PDBePISA<sup>43</sup> v1.37.

**Pull-down assays.** Purified 6 $\times$ His-tagged BoNT/A-RBD and untagged SV2C-LD were centrifuged separately at 16,000g for 15 min at 4 °C. The proteins were then combined and incubated overnight at 4 °C in pull-down buffer (20 mM Tris, pH 7.8, 150 mM NaCl and 5 mM  $\beta$ -mercaptoethanol). The interactions of wild-type BoNT/A-RBD and mutant BoNT/A-RBD with wild-type SV2C-LD were analysed at 5  $\mu$ M and 40  $\mu$ M, respectively. The binding of wild-type SV2C-LD and mutant SV2C-LD to wild-type BoNT/A-RBD was analysed at 10  $\mu$ M and 5  $\mu$ M, respectively. After incubation, the samples were centrifuged, and the supernatants were incubated with ~50  $\mu$ l Ni Sepharose 6 Fast Flow resin (GE Healthcare) for

1 h. The Ni Sepharose resin was washed four times with 2 ml ice-cold pull-down buffer. Proteins were eluted in 100  $\mu$ l pull-down buffer supplemented with 350 mM imidazole. The eluted solutions were analysed by SDS–PAGE.

Peptide competition experiments were performed in the same way except that 50  $\mu$ l 10 mM peptide suspension in pull-down buffer was mixed with the partner protein domain and incubated on ice for 5 min before adding the other protein domain. Final concentrations were 5  $\mu$ M BoNT/A-RBD, 20  $\mu$ M SV2C-LD and 2 mM peptide. Peptides were ordered from Charité. The peptide sequences were as follows: SV2C-A3, NH<sub>2</sub>-DSEFKNCSFFHNK-COOH; BoNT/A-A2 ( $\beta$ -sheet), NH<sub>2</sub>-RGSVMTTNIYLNSS-COOH; and BoNT/A-A5 ( $\beta$ -strand), NH<sub>2</sub>-RGSVMTTN-COOH.

**Protein labelling.** The accessible cysteine residues of SV2C-LD were covalently labelled overnight with fluorescein maleimide (Molecular Probes) in 50 mM Tris, pH 7.5, 150 mM NaCl and 1 mM Tris(2-carboxyethyl)phosphine, using a 20-fold excess of dye. Successful labelling and removal of free fluorescein by desalting (illustra NAP-5 columns, GE Healthcare) was confirmed by SDS–PAGE. The protein:fluorescein ratio was determined from the absorption at 280 nm and 493 nm. The absorption at 280 nm was corrected for the absorption of fluorescein to calculate the protein concentration. The labelling efficiency was approximately 50%.

**K<sub>d</sub> determination.** The binding affinities of SV2C-LD to BoNT/A-RBD (wild-type and mutants) were determined in 50 mM Tris, pH 7.5, and 150 mM NaCl by fluorescence anisotropy titration using 100 nM fluorescein-labelled SV2C-LD (SV2C-LD-F). The anisotropy signal was measured in a Cary Eclipse Fluorescence Spectrophotometer (Agilent Technologies) equipped with automated polarizers ( $\lambda_{\text{ex}}$  480 nm,  $\lambda_{\text{em}}$  520 nm, slit 10 nm, 20 °C, 5 s signal acquisition,  $g = 1.4495$ ). Each data point was averaged from five measurements. To account for the reduced quantum efficiency of fluorescein at pH 5 (20 mM citrate and 150 mM NaCl), the slits were opened to 20 nm. The K<sub>d</sub> values were determined by fitting the data to a one-site-binding model using Origin 7 (OriginLab). The binding affinity of unlabelled SV2C-LD and its mutants was determined by displacing SV2C-LD-F from the complex. The coupled equations for competitive binding were numerically fitted with DataFitter Software (D. Veprintsev, unpublished). The affinities of SV2C-LD mutants were independently confirmed by determining the affinity from the apparent K<sub>d</sub> of the BoNT/A-RBD–SV2C-LD-F complex in the presence of unlabelled SV2C-LD<sup>44</sup>.

**Animals and cells.** All animal experiments were performed in compliance with the guidelines for the welfare of experimental animals issued by the federal government of The Netherlands. All animal experiments were approved by the Animal Ethical Review Committee (DEC) of Utrecht University. HEK293T cells were obtained from the American Type Culture Collection and tested for mycoplasma.

**DNA constructs and antibodies for functional studies.** The SV2C–Flag construct was generated by PCR using the human full-length cDNA (Origene) as a template. The insert was cloned into a GW1-CMV vector, and the sequence was verified. The pGW2-mRFP construct was generated as described<sup>45</sup>. The following primary antibodies were used: goat anti-SV2C (1:400, Santa Cruz Biotechnology, clone P-20, order no. sc-11946), rabbit anti-VGAT (1:400, Synaptic Systems, order no. 131003), rabbit anti-Flag, (1:1,000, Sigma, order no. F7425), mouse anti-His (1:400, NeuroMab, clone N144/14, order no. 75-169) and mouse anti-tubulin- $\beta$ 3 (TUBB3, 1:400, Sigma, order no. PRB-435P). The following secondary antibodies were used: Alexa Fluor 488-conjugated goat anti-mouse antibodies (order no. A11029); Alexa Fluor 568-conjugated goat anti-rabbit antibodies (order no. A11036); Alexa Fluor 647-conjugated goat anti-rabbit antibodies (order no. A21245) and Alexa Fluor 647-conjugated goat anti-mouse antibodies (order no. A21236); Alexa Fluor 488 (order no. A11055)- and Cy5 (order no. 705-175-147)-conjugated donkey anti-goat antibodies; and Cy3-conjugated donkey anti-rabbit antibodies (order no. 711-165-152). Alexa Fluor-conjugated antibodies were purchased from Invitrogen, and cyanine-dye-conjugated antibodies were purchased from Jackson ImmunoResearch. All secondary antibodies were used 1:400.

**Blocking of BoNT/A-RBD uptake by neurons.** Primary striatal cultures were prepared from embryonic day 18 rat brains, as described for hippocampal neurons<sup>26</sup>. Cells were plated on coverslips coated with 35  $\mu$ g ml<sup>-1</sup> poly-L-lysine (Sigma) and 5  $\mu$ g ml<sup>-1</sup> laminin (Roche) at a density of 100,000 cells per well. Striatal cultures were grown in neurobasal medium (NB, Invitrogen) supplemented with B27 (Invitrogen), 0.5 mM glutamine (Invitrogen), 12.5  $\mu$ M glutamate (Sigma) and penicillin/streptomycin mix (Invitrogen) at 37 °C in 5% CO<sub>2</sub>.

At 18 days *in vitro* (DIV), cultured striatal neurons were stimulated using a high K<sup>+</sup> buffer (70 mM KCl, 51.5 mM NaCl, 25 mM HEPES, 30 mM glucose, 2 mM CaCl<sub>2</sub> and 2 mM MgCl<sub>2</sub>; pH 7.4) with or without the GST–BoNT/A-A2 peptide (5  $\mu$ M) or the GST control (5  $\mu$ M), or neurons were treated with a control buffer (2.5 mM KCl, 119 mM NaCl, 25 mM HEPES, 30 mM glucose, 2 mM CaCl<sub>2</sub> and 2 mM MgCl<sub>2</sub>; pH 7.4), for 15 min at 37 °C. The neurons were incubated with GFP–BoNT/A-RBD (200 nM) or GFP only (200 nM) for 10 min at 37 °C. The cells were

then washed with high  $K^+$  or control buffer solution and fixed with 4% paraformaldehyde/4% sucrose in PBS for 10 min.

**Blocking of BoNT/A-RBD binding in HEK293T cells.** HEK293T cells were cultured in DMEM/Ham's F-10 (50/50) medium supplemented with 10% FCS and 1% penicillin/streptomycin at 37 °C in 5%  $CO_2$  and were transfected with pGW2-mRFP and SV2C-Flag, or pGW2-mRFP only, using the FuGENE 6 reagent (Roche). After 24 h, cells were incubated with fresh culture medium containing GST-BoNT/A-A2 or the GST control for 15 min at 37 °C. Next, cells were incubated with BoNT/A-RBD (100 nM) for 10 min at 37 °C, washed with fresh culture medium and fixed with 4% paraformaldehyde/4% sucrose in PBS for 10 min.

**Immunohistochemistry, confocal imaging and image analysis.** After fixation, cells were washed three times with PBS for 5 min and incubated with the indicated primary antibodies in GDB buffer (0.2% BSA, 0.8 M NaCl, 0.5% Triton X-100 and 30 mM phosphate buffer; pH 7.4) overnight at 4 °C. Neurons were then washed three times for 5 min with PBS, incubated with secondary antibodies in GDB buffer for 1 h at room temperature and washed again three times for 5 min with PBS. Coverslips were mounted using VECTASHIELD mounting medium (Vector Laboratories).

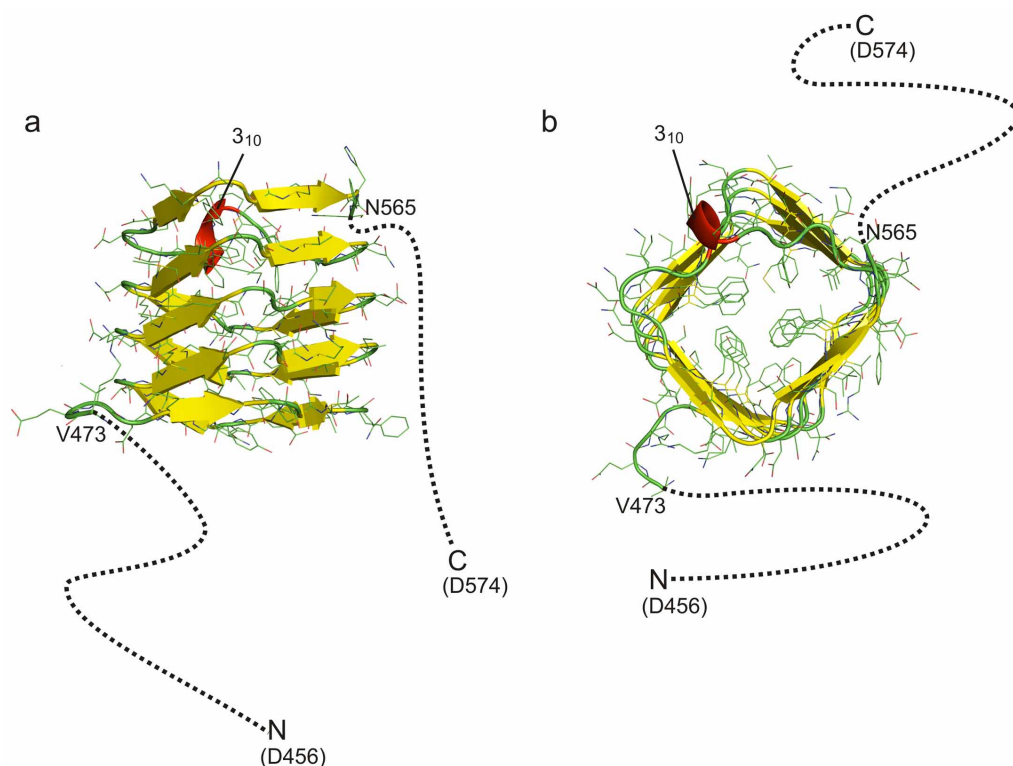
Confocal images were acquired using an LSM 700 confocal laser-scanning microscope (Zeiss) using a 40 $\times$  (1.3 NA) oil objective. Each image was acquired from a z-series of 5 images, each averaged twice, and was chosen to cover the entire region of interest from top to bottom using a 0.47- $\mu$ m interval. The resultant z-stack was 'flattened' into a single image using maximum intensity projection. All image acquisition settings were kept constant throughout the experiments.

Image analysis was performed using ImageJ software (National Institutes of Health). For analysis of the amount of synaptic GFP-BoNT/A-RBD uptake by cultured striatal neurons, the ImageJ plug-in 'Puncta Analyzer' was used to automatically count the number of VGAT- and GFP-BoNT/A-RBD positive puncta. Threshold values were held constant throughout the experiments. To exclude nonspecific background signal, a filter excluding puncta smaller than 0.6  $\mu$ m<sup>2</sup> and larger than 6.3  $\mu$ m<sup>2</sup> was applied. The number of co-localizing VGAT and GFP-BoNT/A-RBD puncta was normalized to the total number of VGAT-positive puncta to correct for changes in neuron density per image. For analysis of the amount of BoNT/A-RBD binding in HEK293T cells, the total area of BoNT/A-RBD was normalized to the total area of mRFP-transfected HEK293T cells above an assigned threshold. Threshold values were kept constant throughout the experiments. Thresholded images were processed using a median filter to remove noise.

**Data representation and statistics.** The number of cells analysed was based on previous data sets<sup>27</sup>. For all experiments, *N* (number of replicates) represents biological replicates. In all bar graphs, the data are presented as the mean values  $\pm$  s.e.m.

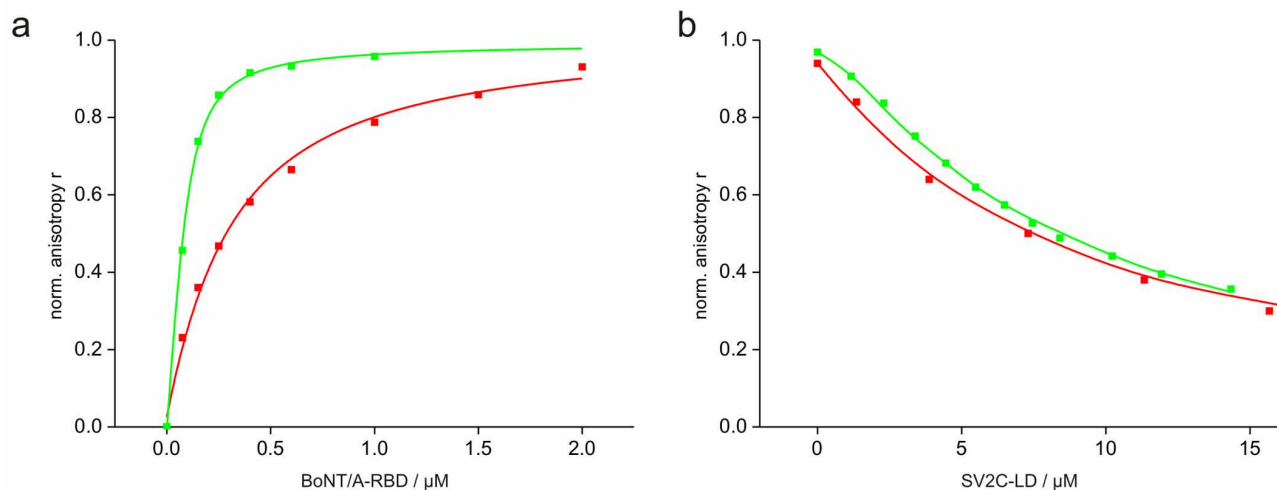
Statistical analysis was performed using SPSS Statistics v20 (IBM). The non-parametric Mann-Whitney *U* test was used to test for statistical significance. *P* values below 0.05 are considered significant (\**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001; NS, not significant). All tests were performed two-sided. The data samples were tested for normal distribution with the Kolmogorov-Smirnov test and for heterogeneity of variance with Bartlett's test.

28. Benoit, R. M., Wilhelm, R. N., Scherer-Becker, D. & Ostermeier, C. An improved method for fast, robust, and seamless integration of DNA fragments into multiple plasmids. *Protein Expr. Purif.* **45**, 66–71 (2006).
29. Olieric, N. et al. Automated seamless DNA co-transformation cloning with direct expression vectors applying positive or negative insert selection. *BMC Biotechnol.* **10**, 56 (2010).
30. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
31. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
32. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
33. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
34. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
36. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
37. DeLano, W. L. The PyMOL Molecular Graphics System. (DeLano Scientific, 2012).
38. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA* **98**, 10037–10041 (2001).
39. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60**, 2256–2268 (2004).
40. Hutchinson, E. G. & Thornton, J. M. PROMOTIF—a program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212–220 (1996).
41. Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**, 363–371 (2009).
42. Heller, H., Schaefer, M. & Schulten, K. Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid crystal phase. *J. Phys. Chem.* **97**, 8343–8360 (1993).
43. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
44. Zhang, Y. L. & Zhang, Z. Y. Low-affinity binding determined by titration calorimetry using a high-affinity coupling ligand: a thermodynamic study of ligand binding to protein tyrosine phosphatase 1B. *Anal. Biochem.* **261**, 139–148 (1998).
45. Kapitein, L. C., Yau, K. W. & Hoogenraad, C. C. Microtubule dynamics in dendritic spines. *Methods Cell Biol.* **97**, 111–132 (2010).



**Extended Data Figure 1 | Structure of the SV2C luminal domain.** Side view (a) and top view (b) of SV2C-LD chain D of the complex structure. One full turn of the  $\beta$ -helix comprises 20 amino acids. The central hydrophobic core of the  $\beta$ -helix is mostly formed by stacked, slightly tilted phenylalanine residues. Cartoon representation: the helix,  $\beta$ -strands and loops are shown in red,

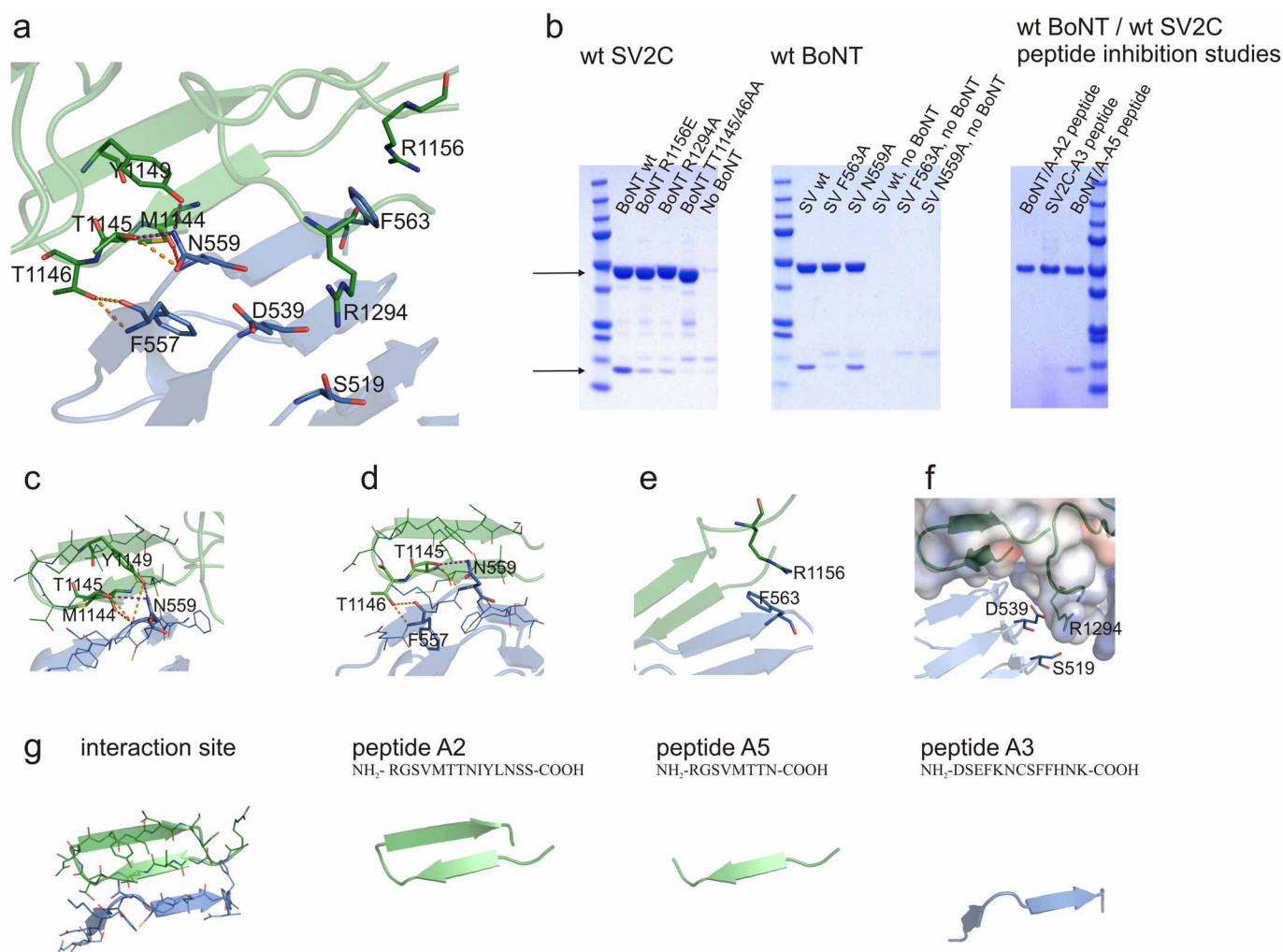
yellow and green, respectively. The side chains are shown as lines in green and in atom colours. The 3<sub>10</sub>-helix is indicated. The flexible N- and C-terminal regions that were not visible in the structure are schematically indicated as dotted lines.



**Extended Data Figure 2 | Binding of BoNT/A-RBD to SV2C-LD is reduced on acidification.** Normalized fluorescence anisotropy titration of BoNT/A-RBD with labelled SV2C-LD (**a**) and displacement with unlabelled

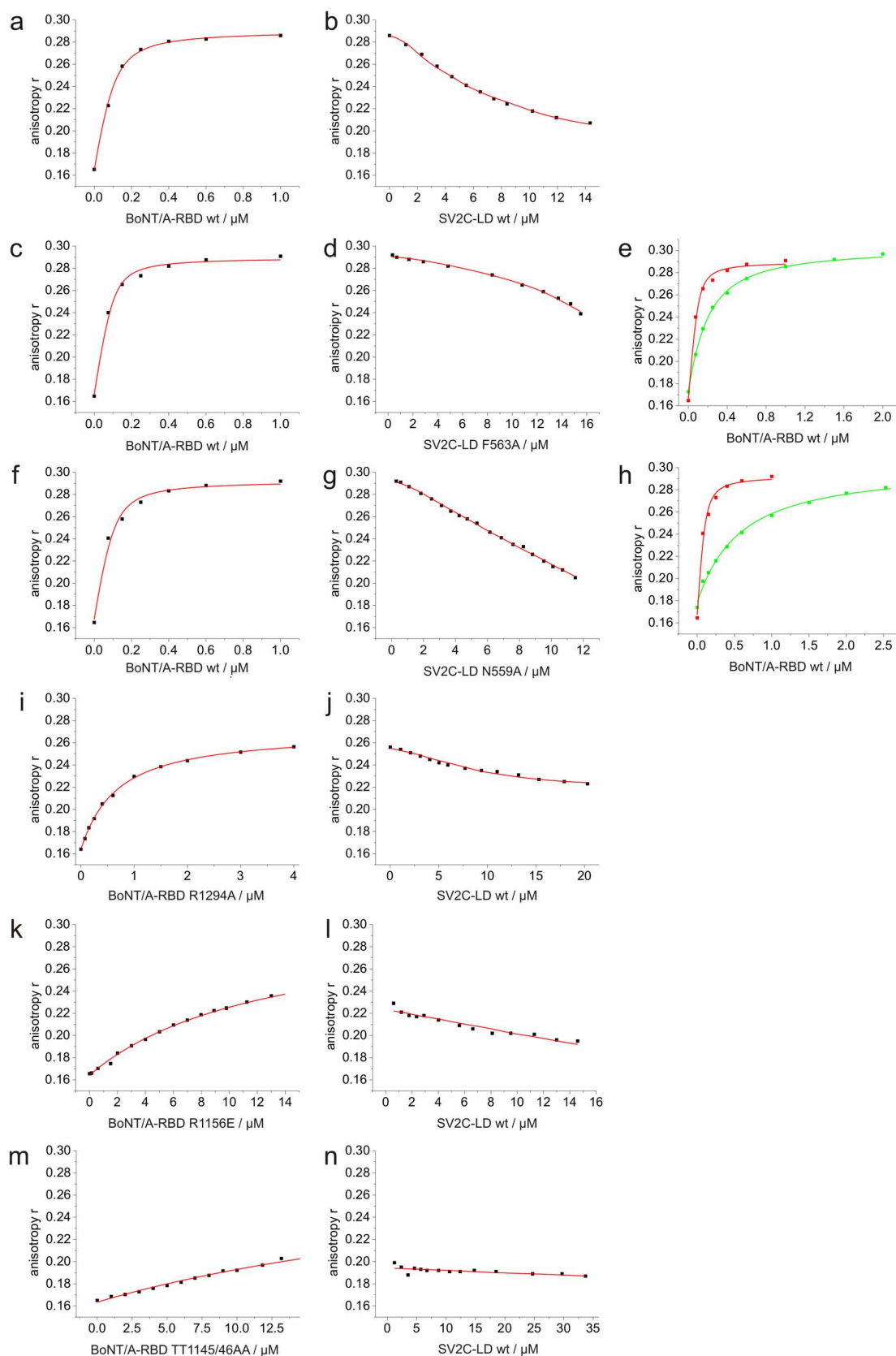
SV2C-LD at pH 7.5 (green) and pH 5 (red) (**b**). The affinity of SV2C-LD for BoNT/A-RBD at pH 5 is reduced by a factor of  $\sim 5$ . For values, see Extended Data Table 3.





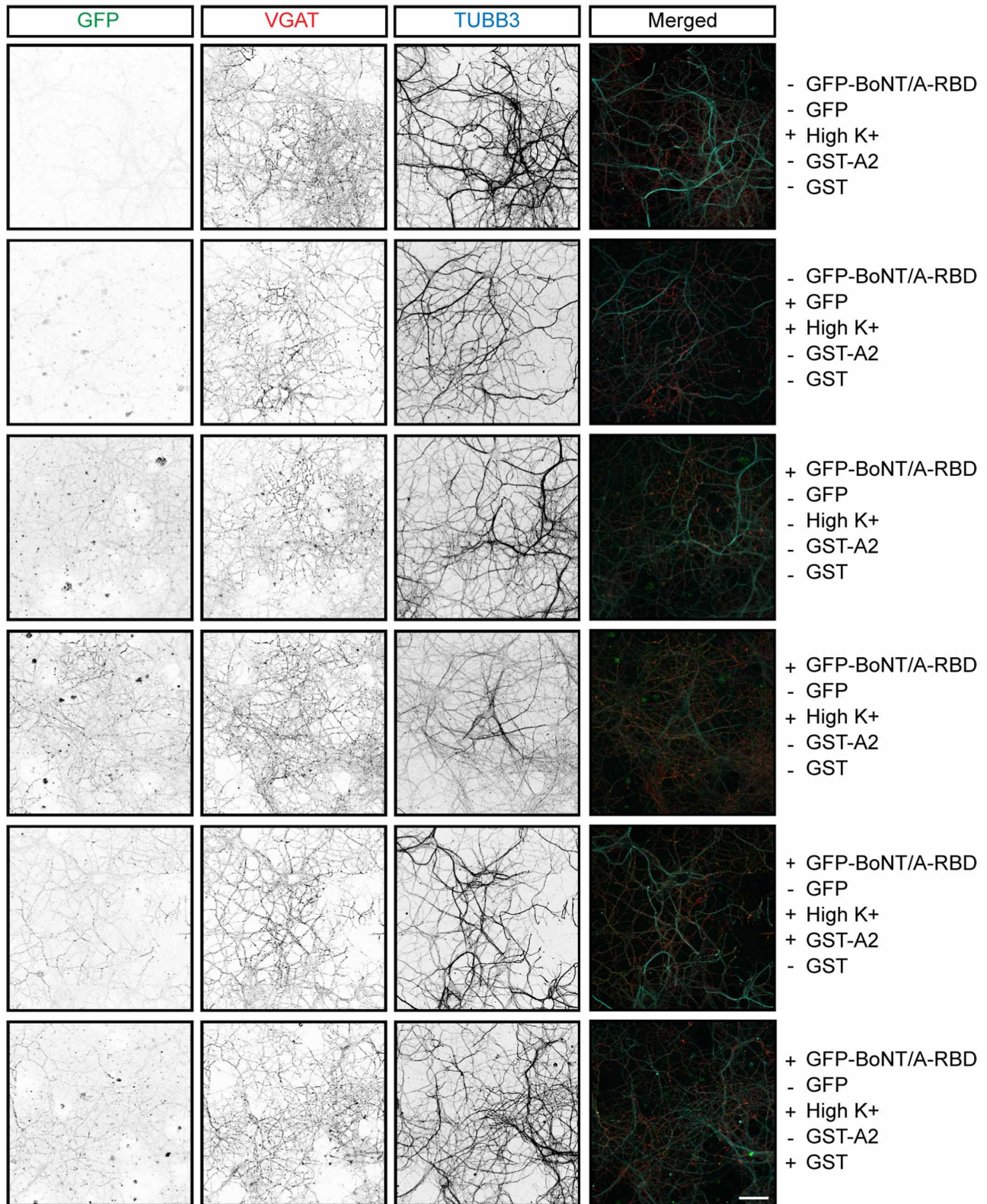
**Extended Data Figure 3 | Interaction between SV2C-LD and BoNT/A-RBD.** **a**, Overview of the prominent interactions that were analysed by site-directed mutagenesis. The colour code in Fig. 1 is used. **b**, SDS-PAGE analysis of the pull-down assays. The 6×His-tagged BoNT/A domain (~50 kDa) and the untagged SV2C domain (~15 kDa) are indicated by arrows. **c–f**, Close-up views of specific interactions. **c**, The hydrogen bonds of N559. **d**, The hydrogen bonds of T1145/T1146. **e**, The cation–π stacking interaction between BoNT/A-RBD R1156 and SV2C-LD F563. **f**, The putative long-range electrostatic interactions of BoNT/A-RBD R1294. R1294 is not

defined in the electron density of the complex structure and hence does not participate in hydrogen bonds or salt bridges. Nevertheless, mutagenesis of R1294 to alanine strongly reduces the binding of BoNT/A-RBD to SV2C-LD. We speculate that long-range electrostatic interactions between the positively charged BoNT/A-RBD arginine (depicted as a surface coloured according to the electrostatic potential) and the negatively charged regions in SV2C-LD have a role in complex formation. **g**, Sequences and schematic representations of the peptides that were used for the complex inhibition studies.



**Extended Data Figure 4 |  $K_d$  determination of wild-type and mutant BoNT/A-RBD and SV2C-LD proteins.** The affinities of wild-type and mutant SV2C-LD for BoNT/A-RBD were determined by fluorescence anisotropy titration of labelled SV2C-LD (a, c, f) and subsequent displacement with unlabelled SV2C-LD (b), SV2C-LD F563A (d) or SV2C-LD N559A (g). Alternatively, the affinities of SV2C-LD mutants were calculated from the

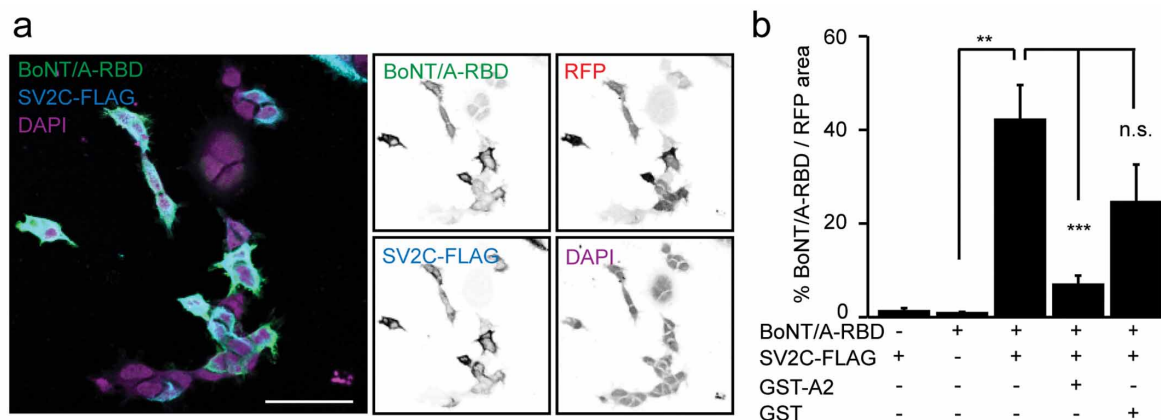
apparent  $K_d$  of labelled SV2C-LD in the presence (green) or absence (red) of SV2C-LD F563A (22.6  $\mu\text{M}$ , e) or SV2C-LD N559A (18.2  $\mu\text{M}$ , h). The affinities of BoNT/A-RBD R1294A (i, j), BoNT/A-RBD R1156E (k, l) and BoNT/A-RBD T1145A/T1146A (m, n) for SV2C-LD were determined accordingly by anisotropy displacement titrations. For values, see Extended Data Table 3.



**Extended Data Figure 5 | The BoNT/A-A2 peptide inhibits the internalization of BoNT/A-RBD by striatal neurons.** Representative images of GFP-BoNT/A-RBD uptake by cultured striatal neurons (DIV18). Neurons were pre-incubated with the GST-BoNT/A-A2 or the GST control (5  $\mu$ M, 15 min) in high K<sup>+</sup> or control buffer and treated with

GFP-BoNT/A-RBD or GFP only (200 nM, 10 min) and stained for the neuronal marker tubulin- $\beta$ 3 (TUBB3) and endogenous VGAT to label presynaptic terminals. Scale bar, 50  $\mu$ m. Representative images are from a total of 20 images per group,  $N = 2$  independent experiments.





**Extended Data Figure 6 | The BoNT/A-A2 peptide inhibits the binding of BoNT/A-RBD in HEK293T cells.** **a**, A typical example of BoNT/A-RBD (green) binding to SV2C-Flag (blue) in HEK293T cells. Cells were transfected with SV2C-Flag and mRFP (red) to highlight transfected cells, fixed and stained using Flag- and His-tagged antibodies. DAPI (4',6-diamidino-2-phenylindole, purple) was used for visualizing all cell nuclei. Scale bar, 50  $\mu$ m. **b**, Quantification of BoNT/A-RBD binding in SV2C-Flag expressing HEK293T cells. Cells were transfected with SV2C-Flag and mRFP, or mRFP only, and incubated with GST-BoNT/A-A2 peptide or GST control

(5  $\mu$ M, 15 min) and treated without (control) or with BoNT/A-RBD (100 nM, 10 min). The total area of BoNT/A-RBD was normalized to the total area of mRFP-transfected cells. No BoNT/A-RBD:  $1.24 \pm 0.64\%$ ,  $n = 8$ ; no SV2C-Flag:  $0.82 \pm 0.64\%$ ,  $n = 5$ ; no GST-BoNT/A-A2:  $42.13 \pm 7.44\%$ ,  $n = 9$ ; 5  $\mu$ M GST-BoNT/A-A2:  $6.92 \pm 1.93\%$ ,  $n = 9$ ; 5  $\mu$ M GST:  $24.55 \pm 8.01\%$ ,  $n = 9$ ; Mann-Whitney  $U$  test, \*\*\*  $P < 0.001$ , \*  $P < 0.05$ , NS not significant,  $n$  = number of images analysed per group,  $N = 2$  independent experiments. All tests were performed two-sided. Data are presented as the mean  $\pm$  s.e.m.



**Extended Data Table 1 | Data collection and refinement statistics****Extended Data Table 1** Data collection and refinement statistics

	BoNT/A-RBD / SV2C-LD complex
<b>Data collection</b>	
Space group	C 1 2 1 (twin operator -h,-k, l)
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	115.44, 105.26, 127.96
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90.02, 90
Resolution (Å)	20-2.3 (2.4-2.3)*
<i>R</i> <sub>meas</sub>	21.3 (102.8)
<i>CC</i> <sub>1/2</sub>	99.1 (68.4)
<i>I</i> / $\sigma$ <i>I</i>	8.3 (2.1)
Completeness (%)	99.8 (99.9)
Redundancy	6.8 (6.7)
<b>Refinement</b>	
Resolution (Å)	20-2.3
No. reflections	67843
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.235/0.269
No. atoms	
Protein	9318
Ligand/ion	8
Water	82
B-factors	
Protein	38.9
Ligand/ion	29.1
Water	20.0
R.m.s deviations	
Bond lengths (Å)	0.008
Bond angles (°)	1.079

\* The highest resolution shell is shown in parentheses.

Extended Data Table 2 | SV2C-LD–BoNT/A-RBD hydrogen bonds

Residue (BoNT/A chain A)	Distance (Å)	Residue (SV2C chain D)
Thr 1146 [N]	3.1	Phe 557 [O]
Thr 1146 [OG1]	3.6	Phe 557 [O]
Met 1144 [N]	3.3	Cys 560 [O]
Ser 1142 [N]	3.2	Phe 562 [O]
Thr 1146 [OG1]	3.0	Phe 557 [N]
Met 1144 [O]	3.5	Asn 559 [N]
Thr 1145 [OG1]	3.0	Asn 559 [N]
Thr 1145 [OG1]	3.5	Asn 559 [ND2]
Tyr 1149 [OH]	2.8	Asn 559 [ND2]
Met 1144 [O]	3.3	Cys 560 [N]
Ser 1142 [O]	2.8	Phe 562 [N]

The table summarizes the hydrogen bonds of the interface between chain A and chain D.

Extended Data Table 3 |  $K_d$  values of the interaction of wild-type and mutant proteins

BoNT/A-RBD variant	wt	wt	wt	wt	R1294A	R1156E	T1145/6AA
SV2C-LD variant	wt	wt, pH5	F563A	F559A	wt	wt	Wt
$K_d$ labeled [nM]	28±3	260±40	16±5	23±9	610±30	12,800±500	37,000±2,000
$K_d$ unlabeled [nM] <sup>a</sup>	260±20	1,400±100	1,700±400	400±20	1,000±200	≥7,000	≥25,000
$K_d$ app [nM]			143±7	500±70			
$K_d$ unlabeled [nM] <sup>b</sup>			2,800±1000	850±150			
Pull-down [interaction]	strong		moderate	strong	moderate	moderate	no

<sup>a</sup>Displacement with unlabelled protein. <sup>b</sup>In the presence of the unlabelled mutant SV2C-LD. Please note that the affinity of SV2C-LD for BoNT/A-RBD R1156E and T1145A/T1146A is reduced to an extent that allows only determination of a lower boundary for the affinity of unlabelled SV2C-LD (Extended Data Fig. 4l, n). However, these values are consistent with the reduced binding affinity of labelled SV2C-LD and the pull-down experiments.

# Coupled GTPase and remodelling ATPase activities form a checkpoint for ribosome export

Yoshitaka Matsuo<sup>1</sup>, Sander Granneman<sup>2,3\*</sup>, Matthias Thoms<sup>1\*</sup>, Rizos-Georgios Manikas<sup>1</sup>, David Tollervey<sup>2</sup> & Ed Hurt<sup>1</sup>

Eukaryotic ribosomes are assembled by a complex pathway that extends from the nucleolus to the cytoplasm and is powered by many energy-consuming enzymes<sup>1–3</sup>. Nuclear export is a key, irreversible step in pre-ribosome maturation<sup>4–8</sup>, but mechanisms underlying the timely acquisition of export competence remain poorly understood. Here we show that a conserved *Saccharomyces cerevisiae* GTPase Nug2 (also known as Nog2, and as NGP-1, GNL2 or nucleostemin 2 in human<sup>9</sup>) has a key role in the timing of export competence. Nug2 binds the inter-subunit face of maturing, nucleoplasmic pre-60S particles, and the location clashes with the position of Nmd3, a key pre-60S export adaptor<sup>10</sup>. Nug2 and Nmd3 are not present on the same pre-60S particles, with Nug2 binding before Nmd3. Depletion of Nug2 causes premature Nmd3 binding to the pre-60S particles, whereas mutations in the G-domain of Nug2 block Nmd3 recruitment, resulting in severe 60S export defects. Two pre-60S remodelling factors, the Rea1 ATPase and its co-substrate Rsa4, are present on Nug2-associated particles, and both show synthetic lethal interactions with *nug2* mutants. Release of Nug2 from pre-60S particles requires both its K<sup>+</sup>-dependent GTPase activity and the remodelling ATPase activity of Rea1. We conclude that Nug2 is a regulatory GTPase that monitors pre-60S maturation, with release from its placeholder site linked to recruitment of the nuclear export machinery.

The conserved GTPase Nug2 (Extended Data Fig. 1) is associated with several pre-60S particles located in the nucleoplasm, but was not detected on particles with a known cytoplasmic location (see also Extended Data Fig. 2). The bacterial homologue of Nug2, YlqF, binds directly to ribosomal RNA<sup>11</sup>, and we therefore used the ultraviolet cross-linking and analysis of complementary DNA (CRAC) method to localize the binding site for yeast Nug2 within the pre-60S particle<sup>12</sup>. Direct contacts for Nug2 were identified only with the 25S rRNA, at sites in helices H38, H69, H71, H80, H81–83, H84–86, H89, H91–92 and H93 (Fig. 1a, c). Yeast three-hybrid analyses confirmed interactions between Nug2 and these rRNA helices (Fig. 1b). Mapping the major rRNA crosslink sites of Nug2 onto the 60S subunit structure (Fig. 1d) showed a distinct cluster on the inter-subunit joining surface<sup>13</sup>. Nug2-binding sites overlap with regions occupied by the export factor Nmd3 in cryo-electron microscopy<sup>10</sup>. CRAC was therefore applied to Nmd3 to identify its binding sites more precisely, which were found to lie in H38, H69 and H89 of 25S rRNA (Fig. 1a, c, e). Notably, the Nug2- and Nmd3-binding sites overlapped in H38, H69 and H89 (Fig. 1c, f), suggesting that the binding of these two proteins is mutually exclusive. To test this, pre-60S particles were purified with tagged Nug2 and shown to lack detectable Nmd3, and vice versa (Extended Data Fig. 2). Nmd3 is an essential nuclear export factor that recruits the export receptor Crm1 to the nascent 60S subunits<sup>14,15</sup>. These observations suggested that Nug2 acts as a 'placeholder' to prevent premature recruitment of Nmd3 to earlier, export-incompetent pre-60S particles.

Like other GTP-binding proteins, Nug2 has characteristic G1, G3 and G4 motifs in its G-domain (Fig. 2a and Extended Data Fig. 1),

suggesting that GTP-binding or hydrolysis<sup>16</sup> might regulate dynamic interactions between Nug2 and the pre-ribosome. Dominant-negative mutations were previously described in two GTPases involved in ribosome biogenesis, the G1 motif of Lsg1 (Lys349Asn/Arg/Thr)<sup>17</sup> and the G3 motif of Nog1 (Gly224Ala)<sup>18</sup>. Orthologous G1- and G3-motif mutants, *nug2(K328R)* and *nug2(G369A)*, respectively (Fig. 2a and Extended Data Fig. 1), each showed severe growth defect phenotypes (Fig. 2b), and were also dominant-negative when overexpressed in the presence of chromosomal *NUG2* (Fig. 2c). Pre-ribosome analysis by sucrose gradient centrifugation showed that the Nug2(Lys328Arg) and Nug2(Gly369Ala) proteins were efficiently assembled into pre-60S subunits, but induced a 'half-mer' polysome phenotype (in particular for Nug2(Lys328Arg)), characteristic of reduced 60S subunit synthesis (Fig. 2d). The reduced 60S levels were more apparent under low Mg<sup>2+</sup> conditions that cause 80S ribosomes to dissociate into 60S and 40S subunits (Fig. 2d). The *nug2(K328R)* and *nug2(G369A)* strains showed nuclear accumulation of an enhanced green fluorescent protein (eGFP)-containing Rpl25-eGFP reporter, but not Rps3-eGFP, revealing a specific block in pre-60S nuclear export (Fig. 2e). We conclude that mutations in the GTPase domain of Nug2 allow recruitment to the pre-ribosomes, but block nuclear export.

To determine the basis of the defects associated with Nug2(Lys328Arg) and Nug2(Gly369Ala), we assayed *in vitro* guanine-nucleotide-binding activity and GTP hydrolysis. Nug2 from *S. cerevisiae* was unstable when expressed in *Escherichia coli* (data not shown). By contrast, good yields were obtained for wild-type and mutant Nug2 from the eukaryotic thermophile *Chaetomium thermophilum* (ctNug2, ctNug2(Lys339Arg) and ctNug2(Gly380Ala), respectively; Fig. 2f), the thermostable proteins of which have superior biochemical properties<sup>19</sup>. ctNug2 is highly homologous to yeast Nug2 (74% identity; Extended Data Fig. 1), and can complement, albeit not perfectly, a yeast *nug2Δ* mutant (Extended Data Fig. 3). As Nug2 may act as a potassium-dependent GTPase<sup>20</sup>, we tested the cation requirement for GTP hydrolysis. The GTPase activity of ctNug2 was low in NaCl-containing buffer, but was substantially stimulated by KCl (Fig. 2f). By contrast, ctNug2(Lys339Arg) and ctNug2(Gly380Ala) exhibited only background GTPase activity (Fig. 2f). In binding assays, wild-type ctNug2 and ctNug2(Gly380Ala) readily bound the fluorescent nucleotides MANT-GTP or MANT-GDP, whereas ctNug2(Lys339Arg) did not (Fig. 2g). We conclude that ctNug2(Lys339Arg) is defective in GTP binding, whereas ctNug2(Gly380Ala) binds but cannot hydrolyse GTP. This K<sup>+</sup>-stimulated GTPase activity might regulate the interaction of Nug2 with nascent 60S particles.

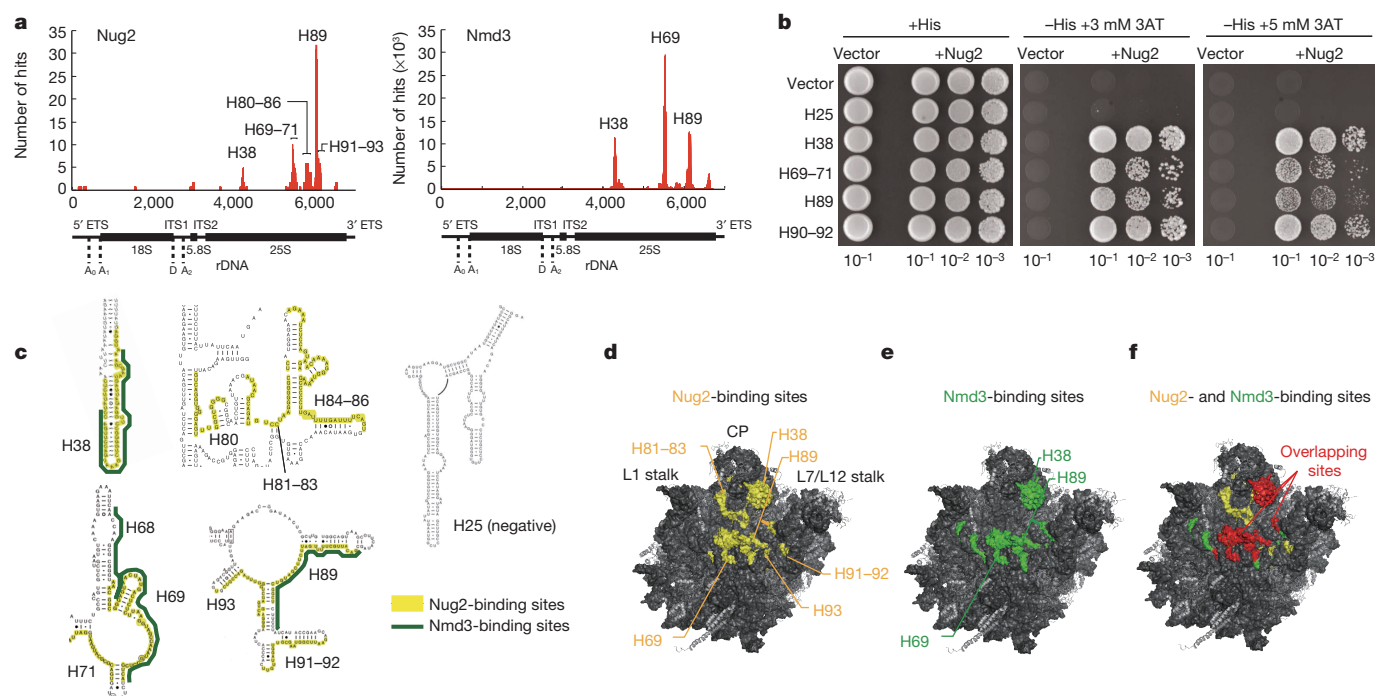
Nug2 is associated with nucleoplasmic pre-60S particles that also carry the Rix1–Ipi1–Ipi3 heterotrimer, the dynein-related AAA-ATPase Rea1, and its co-substrate Rsa4 (Extended Data Fig. 2; see also below and ref. 21). The enzymatic activity of Rea1 is required for the release of Ytm1 (ref. 22) and Rsa4 (ref. 21) and a genetic screen revealed synthetic lethality between the G1-motif mutant *nug2(K328R)* and the mutant alleles *rea1-DTS* and *rsa4-1* (ref. 21) (Fig. 3a). We therefore investigated whether ATP-dependent remodelling of the Rix1 particle by the

<sup>1</sup>Biochemie-Zentrum der Universität Heidelberg, Im Neuenheimer Feld 328, Heidelberg D-69120, Germany. <sup>2</sup>Wellcome Trust Centre for Cell Biology, The University of Edinburgh, Edinburgh EH9 3JR, UK.

<sup>3</sup>Centre for Synthetic and Systems Biology, The University of Edinburgh, Edinburgh EH9 3JD, UK.

\*These authors contributed equally to this work.





**Figure 1 | Nug2 binds to inter-subunit face of the pre-60S subunit clashing with export factor Nmd3.** **a**, CRAC analyses of Nug2 and Nmd3 (performed twice; only sites were considered that were reproducibly found in both data sets). The total number of hits was plotted against the relative location along the rDNA. ETS, external transcribed spacer; ITS, internal transcribed spacer region. A<sub>0</sub>, A<sub>1</sub>, A<sub>2</sub> and D indicate RNA cleavage sites within the pre-rRNA unit. **b**, Yeast three-hybrid analysis revealing interaction between Nug2 and

identified 25S rRNA fragments. Negative control, empty vector and H25. 3AT, 3-amino-1,2,4-triazole. **c**, Nug2- (yellow) and Nmd3- (green) binding sites identified by CRAC and highlighted in the indicated 25S rRNA. **d**, **e**, Mapping of CRAC Nug2- (yellow) and Nmd3- (green) binding sites on the 60S structure (PDB code 3O5H; ref. 13). **f**, Overlapping binding sites (red) of Nug2 (yellow) and Nmd3 (green).

AAA-ATPase activity of Rea1 (ref. 21) is altered in particles containing Nug2(Lys328Arg) or Nug2(Gly369Ala). Pre-60S particles carrying Flag-tagged Rpl3 were affinity purified using a tandem affinity purification (TAP)-tagging technique (Rix1-TAP) via IgG binding and tobacco etch virus (TEV) elution. The pre-60S particles were incubated *in vitro* to allow factor release, and then re-isolated on Flag beads via Rpl3-Flag (Fig. 3b). Consistent with previous data<sup>21</sup>, incubation of the pre-60S particles with ATP in Na<sup>+</sup>-containing buffer resulted in the release of Rsa4 and Rea1, but not Nug2 (Fig. 3c). By contrast, incubation in K<sup>+</sup>-containing buffer caused the ATP-dependent release of Nug2, in addition to Rsa4 and Rea1 (Fig. 3c). Incubation with GTP in Na<sup>+</sup>- or K<sup>+</sup>-containing buffer did not induce the release of biogenesis factors (Fig. 3d). However, neither Nug2(Lys328Arg) nor Nug2(Gly369Ala) could be dissociated from pre-60S particles after ATP treatment in K<sup>+</sup> buffer (Fig. 3e). In the case of Nug2(Lys328Arg) (defective in GTP binding), incubation with ATP in K<sup>+</sup> buffer failed to release Rsa4, whereas pre-60S particles carrying Nug2(Gly369Ala) (defective in GTP hydrolysis) still showed Rsa4 release after ATP treatment (Fig. 3e). Mutation of one of the six ATP-binding protomers of Rea1 (AAA2; *rea1* Lys659Ala) inhibited remodelling, including Nug2 release (Extended Data Fig. 4). These findings indicate that the GTP-binding activity of Nug2 influences the remodelling activity of the Rea1 ATPase, whereas GTP hydrolysis is necessary for the final Nug2 release from the pre-60S subunit.

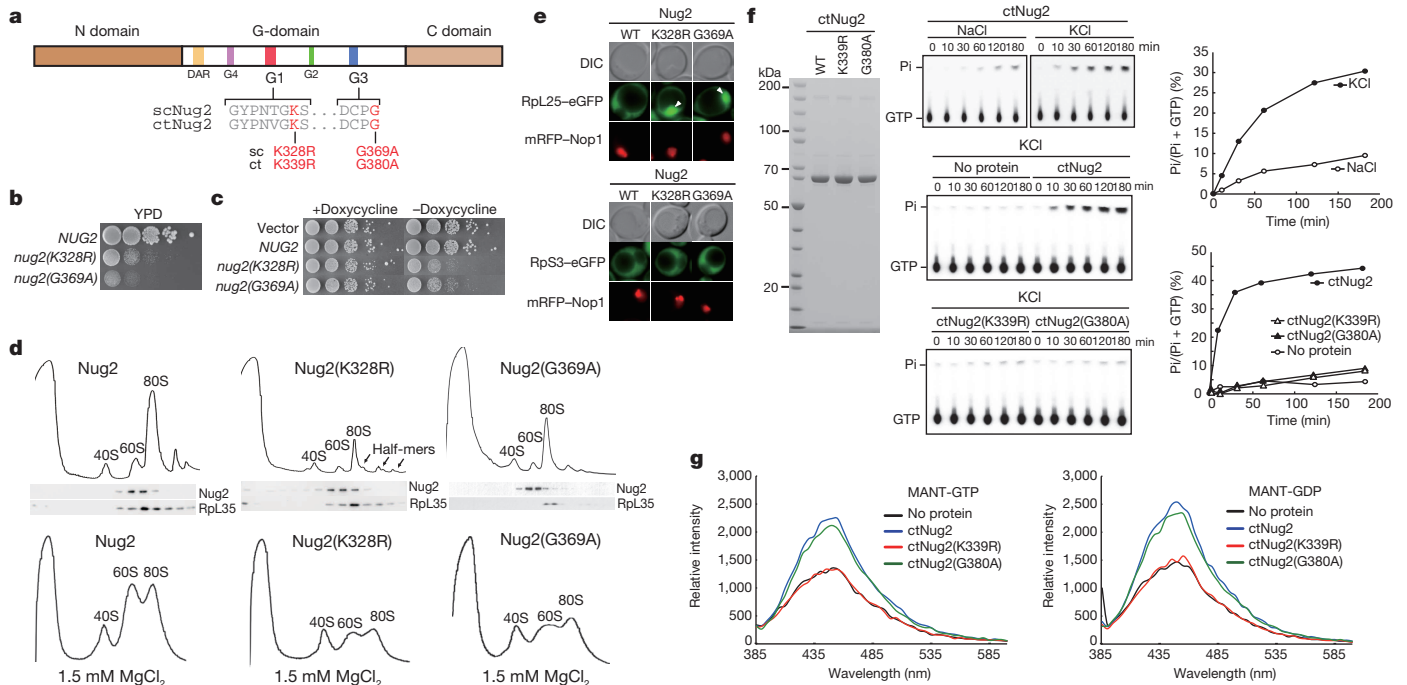
*In vitro*, Rea1-dependent release of Rsa4 and Nug2 required only ATP and K<sup>+</sup> without the addition of GTP, whereas the mutational analyses suggested that GTPase activity is necessary for Nug2 release. These findings suggest that Nug2 on the Rix1 particle might have retained bound GTP during purification (which is possible owing to its low intrinsic GTPase activity). Alternatively, ribosome-associated nucleotide diphosphate kinases can transfer the  $\gamma$ -phosphate from ATP to GDP to generate GTP-loaded GTPases<sup>23,24</sup>.

The pre-60S particles co-purified with Rix1 also contained small amounts of Ytm1 and Erb1 (Fig. 3c), which were previously described as nucleolar co-substrates for Rea1 (ref. 22), and both were released by incubation with ATP in Na<sup>+</sup> or K<sup>+</sup> buffer (Fig. 3c).

To determine the step in 60S subunit biogenesis at which dissociation of Nug2 is disturbed *in vivo*, we affinity-purified different pre-60S particles from Nug2 wild-type and mutant cells using bait proteins that specifically enrich nucleolar, nucleoplasmic or cytoplasmic intermediates (Fig. 4a). The *nug2*(K328R) mutation did not markedly alter the biochemical composition of most pre-60S particles tested. The exception was Arx1-associated particles, which showed a marked depletion of the export adaptor Nmd3 and the cytoplasmic factor Rei1 that stimulates recycling of Arx1 (ref. 25) (Fig. 4a). Nmd3 was also largely absent from Arx1 particles purified from Nug2(Gly369Ala) cells (Fig. 4b).

To test the model that Nug2 depletion allows premature recruitment of Nmd3, we used an auxin-inducible degron system<sup>26</sup>. Nug2 was expressed as a fusion protein (sAid-Nug2-sAid) with two copies of the sAid (small auxin-inducible degron) tag, which is targeted by the F-box E3 ubiquitin ligase TIR1 in the presence of auxin, inducing fast proteasomal degradation<sup>26</sup> (Extended Data Fig. 5). Nmd3 is normally not detected on Rix1-associated particles, but was prematurely recruited to this pre-60S particle after Nug2 depletion (Fig. 4c). Concomitant with Nmd3 association, the recovery of Rea1 and Rsa4 decreased during Nug2 depletion (Fig. 4c). We conclude that Nug2 promotes the stable association of Rsa4 and Rea1 with the Rix1 particles, while blocking premature recruitment of Nmd3.

To address the timing of Nug2 recruitment to pre-60S particles in comparison to Rea1, Rsa4 and the Rix1-Ipi1-Ipi3 complex, we used a combination of affinity purification and immunodepletion. Affinity purification of Nug2-TAP yielded a mixture of different pre-60S particles including Rix1/Rea1 particles. Rix1-Flag immunoprecipitation was used to deplete Rix1/Rea1 particles from this mixture, leaving

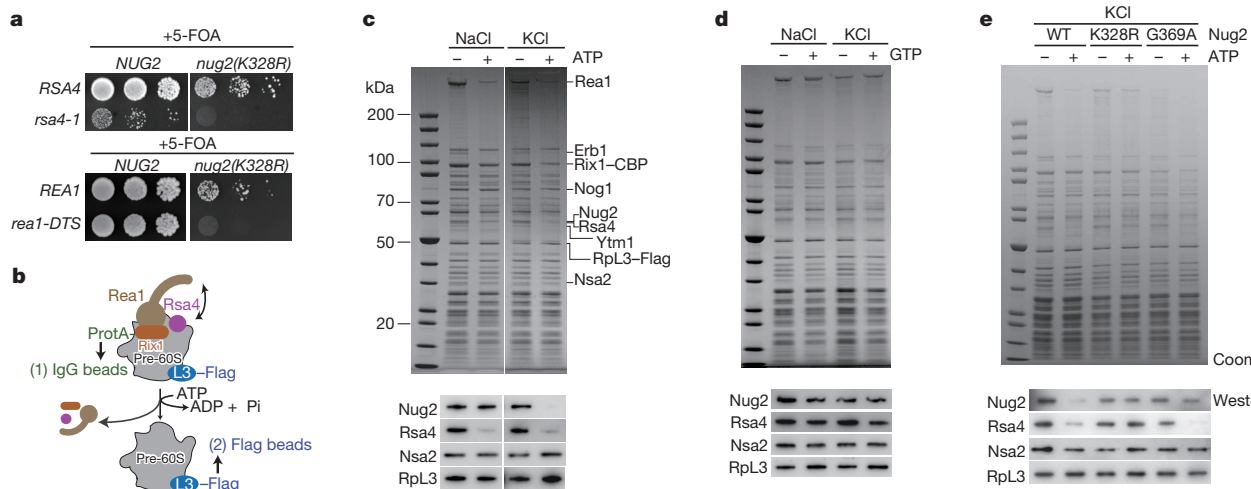


**Figure 2 |  $K^+$ -dependent GTPase activity of Nug2.** **a**, Domain organization of Nug2. C, carboxy; N, amino. **b**, Complementation of *nug2Δ* cells by *NUG2*, *nug2(K328R)* and *nug2(G369A)* on YPD plates. **c**, Repression (+doxycycline) and overexpression (–doxycycline) of *NUG2*, *nug2(K328R)* and *nug2(G369A)* in *NUG2* cells. **d**, Polysomal (10 mM  $MgCl_2$ ; top) and ribosomal (1.5 mM  $MgCl_2$ ; bottom) profiles of *Nug2*, *Nug2(K328R)* and *Nug2(G369A)* cells analysed by sucrose gradient centrifugation. Western analysis of gradient fractions using antibodies against *Nug2* and *Rpl35*. **e**, Subcellular distribution

*Nug2* particles that contained *Rsa4* and several intermediate pre-60S factors including *Nog1*, *Arx1*, *Nug1*, *Nop53*, *Nsa3*, *Rpf2*, *Rlp7* and *Nsa2* (Fig. 4d). However, this *Nug2* particle lacked other (further upstream) pre-60S factors such as *Ytm1*, *Erb1* and *Has1*, suggesting that it corresponds to the precursor particle to which *Nug2* was recruited. These data complement previous findings that *Nug2* is the last ‘B-factor’ to associate with pre-ribosomes after dissociation of *Has1* (ref. 27).

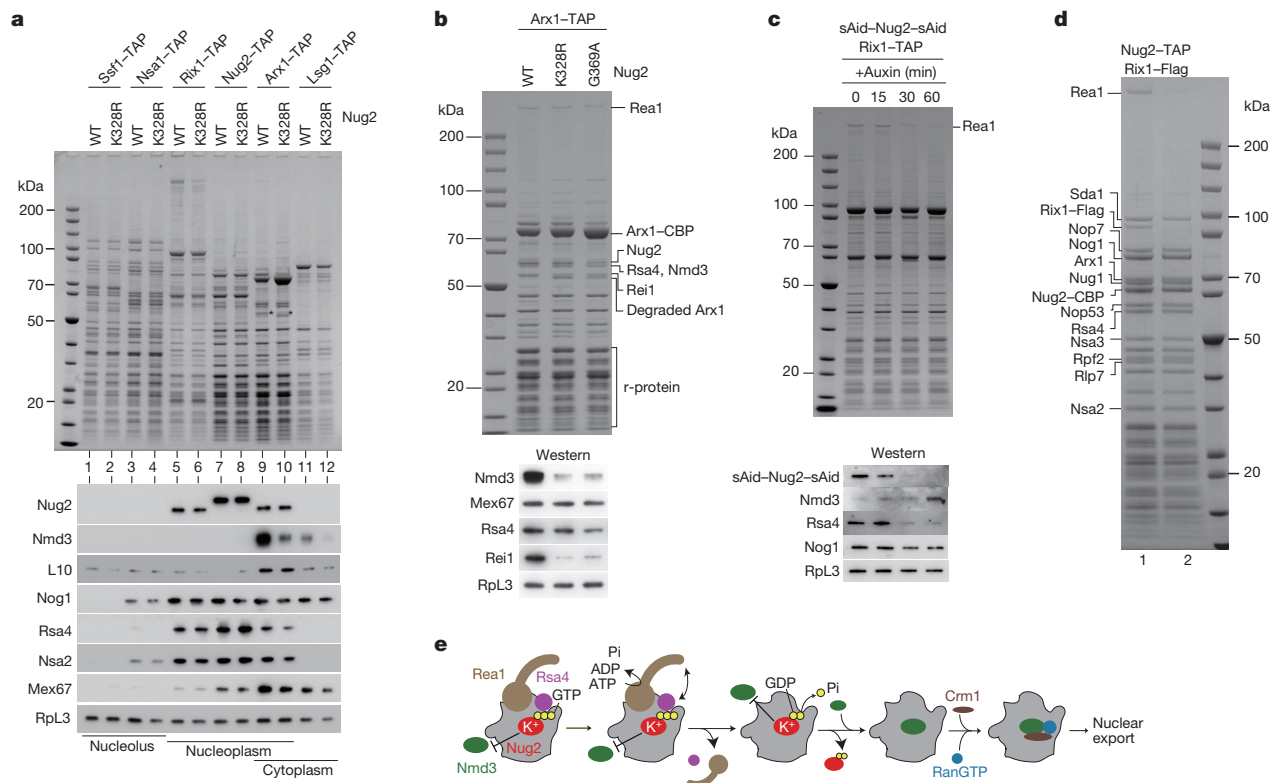
of *Rpl25*–eGFP and *Rps3*–eGFP in *NUG2* and *nug2* mutant cells analysed by fluorescence microscopy. DIC, differential interference contrast; mRFP, monomeric red fluorescent protein. **f**, GTPase activity of purified *ctNug2* (SDS–PAGE; left) analysed by thin-layer chromatography/autoradiography (middle). Ratio of hydrolysed phosphate/total GTP plotted against time (right). **g**, Binding of fluorophores MANT-GTP (left) and MANT-GDP (right) to purified wild-type and mutant *ctNug2*. GTPase and binding assays were performed twice yielding highly reproducible data sets.

As outlined in Fig. 4e, we propose that a previously uncharacterized step in the reorganization of the evolving pre-60S subunit primes it for nuclear export. This involves a regulatory GTPase *Nug2* that overlaps with the binding site for the essential nuclear export adaptor *Nmd3*. As long as intranuclear maturation is incomplete, the pre-60S subunit cannot be exported, because recruitment of this essential export factor is not possible. However, a late nucleoplasmic remodelling step, catalysed



**Figure 3 | *Nug2* release from pre-60S particles requires intrinsic  $K^+$ -dependent GTPase and *Rea1* ATPase activity.** **a**, Synthetic lethality between alleles *rsa4-1* (ref. 21) or *rea1-DTS*<sup>21</sup> and *nug2(K328R)* revealed by growth on 5-fluoroorotic acid (5-FOA). **b–e**, ATP-dependent release of *Rsa4* and *Nug2* from purified pre-60S particles. Scheme of the release assay (**b**) and experimental analyses (**c–e**). Affinity-purified *Rix1* particles carrying wild-type or mutant *Nug2* were incubated with ATP or GTP in NaCl or KCl buffer, before

matured pre-60S particles were re-isolated via *Rpl3*–Flag affinity-purification. Final eluates were analysed by SDS–PAGE and Coomassie staining (top); indicated bands were identified by mass spectrometry and western blotting using the indicated antibodies (bottom) (**c–e**). CBP, calmodulin-binding peptide. All *in vitro* assays were performed at least twice with highly reproducible data sets.



**Figure 4 | Nug2 release from the pre-60S subunit is linked to Nmd3 recruitment.** **a**, Affinity-purification of the indicated TAP-tagged pre-60S factors from *NUG2* or *nug2(K328R)* mutant cells. Asterisks denote position of Reil identified by mass spectrometry. WT, wild type. **b**, Affinity-purification of Arx1-TAP from *NUG2*, *nug2(K328R)* and *nug2(G369A)* cells. r-protein denotes ribosomal proteins. **c**, Affinity-purification of Rix1-TAP from sAid-Nug2-sAid degenron strain after time-dependent auxin treatment.

by the AAA-ATPase Rea1 and its co-factor Rsa4, restructures the pre-60S particle, which could lead to both an rRNA and assembly factor rearrangement. This conformational change could also stimulate the  $K^+$ -dependent GTPase activity of Nug2, thereby triggering its release from the matured pre-60S particles. We suggest that the Nug2 GTPase acts as molecular switch to proofread pre-ribosome maturation and regulate the acquisition of export competence. After this reorganization step, the binding site for Nmd3 becomes accessible on the pre-60S subunit, which further triggers Crm1 and RanGTP recruitment to generate nuclear export competence. Thus, our data indicate coordination between a remodelling AAA-ATPase and a conformation-sensing GTPase.

The human Nug2 orthologue GNL2 is highly expressed in proliferating cells including cancer cells, and is involved in the control of cell cycle progression<sup>28</sup>. The discovery of the role of Nug2 during surveillance of ribosome biogenesis may help to reveal the molecular mechanisms by which nucleostemin family members interconnect the elementary cellular processes of ribosome biogenesis and cell proliferation.

## METHODS SUMMARY

Materials and methods for TAP purification, CRAC analysis, purification of ctNug2, GTPase and guanine nucleotide binding assays are described in detail in the Methods. Yeast strains and plasmids used in this study are described in Extended Data Tables 1 and 2. Adapters used for the CRAC analysis are described in Extended Data Table 3.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 June; accepted 3 October 2013.

Published online 17 November 2013.

- Strunk, B. S. & Karbstein, K. Powering through ribosome assembly. *RNA* **15**, 2083–2104 (2009).

**d**, Affinity-purification of Nug2-TAP particles with (lane 1) or without (lane 2) subsequent Rix1-Flag immunodepletion. **a–d**, SDS-PAGE and Coomassie staining (top) and western blotting using the indicated antibodies (bottom). Protein bands indicated were identified by mass spectrometry. All affinity purifications were performed at least twice, yielding highly reproducible data sets. **e**, Model of pre-60S subunit maturation starting from the Rix1 particle with final Nmd3-Crm1-RanGTP recruitment.

- Staley, J. P. & Woolford, J. L. Jr. Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Curr. Opin. Cell Biol.* **21**, 109–118 (2009).
- Granneman, S. & Baserga, S. J. Ribosome biogenesis: of knobs and RNA processing. *Exp. Cell Res.* **296**, 43–50 (2004).
- Lafontaine, D. L. J. A 'garbage can' for ribosomes: how eukaryotes degrade their ribosomes. *Trends Biochem. Sci.* **35**, 267–277 (2010).
- Warner, J. R. & McIntosh, K. B. How common are extraribosomal functions of ribosomal proteins? *Mol. Cell* **34**, 3–11 (2009).
- Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763–776 (2009).
- Zemp, I. & Kutay, U. Nuclear export and cytoplasmic maturation of ribosomal subunits. *FEBS Lett.* **581**, 2783–2793 (2007).
- Dez, C., Houseley, J. & Tollervey, D. Surveillance of nuclear-restricted pre-ribosomes within a subnuclear region of *Saccharomyces cerevisiae*. *EMBO J.* **25**, 1534–1546 (2006).
- Tsai, R. Y. & Meng, L. Nucleostemin: a latecomer with new tricks. *Int. J. Biochem. Cell Biol.* **41**, 2122–2124 (2009).
- Sengupta, J. *et al.* Characterization of the nuclear export adaptor protein Nmd3 in association with the 60S ribosomal subunit. *J. Cell Biol.* **189**, 1079–1086 (2010).
- Matsuo, Y. *et al.* The GTP-binding protein YlqF participates in the late step of 50 S ribosomal subunit assembly in *Bacillus subtilis*. *J. Biol. Chem.* **281**, 8110–8117 (2006).
- Granneman, S., Kudla, G., Petfalski, E. & Tollervey, D. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl Acad. Sci. USA* **106**, 9613–9618 (2009).
- Ben-Shem, A., Jenner, L., Yusupova, G. & Yusupov, M. Crystal structure of the eukaryotic ribosome. *Science* **330**, 1203–1209 (2010).
- Gadal, O. *et al.* Nuclear export of 60S ribosomal subunits depends on Xpo1p and requires a nuclear export sequence-containing factor, Nmd3p, that associates with the large subunit protein Rpl10p. *Mol. Cell Biol.* **21**, 3405–3415 (2001).
- Ho, J. H., Kallstrom, G. & Johnson, A. W. Nmd3p is a Crm1p-dependent adaptor protein for nuclear export of the large ribosomal subunit. *J. Cell Biol.* **151**, 1057–1066 (2000).
- Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: conserved structure and molecular mechanism. *Nature* **349**, 117–127 (1991).
- Hedges, J., West, M. & Johnson, A. W. Release of the export adaptor, Nmd3p, from the 60S ribosomal subunit requires Rpl10p and the cytoplasmic GTPase Lsg1p. *EMBO J.* **24**, 567–579 (2005).

18. Lapik, Y. R., Misra, J. M., Lau, L. F. & Pestov, D. G. Restricting conformational flexibility of the switch II region creates a dominant-inhibitory phenotype in Obg GTPase Nog1. *Mol. Cell. Biol.* **27**, 7735–7744 (2007).
19. Amlacher, S. *et al.* Insight into structure and assembly of the nuclear pore complex by utilizing the genome of a eukaryotic thermophile. *Cell* **146**, 277–289 (2011).
20. Ash, M. R., Maher, M. J., Mitchell Guss, J. & Jormakka, M. The cation-dependent G-proteins: in a class of their own. *FEBS Lett.* **586**, 2218–2224 (2012).
21. Ulbrich, C. *et al.* Mechanochemical removal of ribosome biogenesis factors from nascent 60S ribosomal subunits. *Cell* **138**, 911–922 (2009).
22. Baßler, J. *et al.* The AAA-ATPase Rea1 drives removal of biogenesis factors during multiple stages of 60S ribosome assembly. *Mol. Cell* **38**, 712–721 (2010).
23. Kikkawa, S. *et al.* Conversion of GDP into GTP by nucleoside diphosphate kinase on the GTP-binding proteins. *J. Biol. Chem.* **265**, 21536–21540 (1990).
24. Wertheimer, A. M. & Kaulenas, M. S. GDP kinase activity associated with salt-washed ribosomes. *Biochem. Biophys. Res. Commun.* **78**, 565–571 (1977).
25. Hung, N. J. & Johnson, A. W. Nuclear recycling of the pre-60S ribosomal subunit-associated factor Arx1 depends on Rei1 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **26**, 3718–3727 (2006).
26. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods* **6**, 917–922 (2009).
27. Dembowski, J. A., Kuo, B. & Woolford, J. L. Jr. Has1 regulates consecutive maturation and processing steps for assembly of 60S ribosomal subunits. *Nucleic Acids Res.* **41**, 7889–7904 (2013).
28. Chennupati, V. *et al.* Signals and pathways regulating nucleolar retention of novel putative nucleolar GTPase NGP-1 (GNL-2). *Biochemistry* **50**, 4521–4536 (2011).

**Acknowledgements** We thank M. Remacha, M. Fromont-Racine, A. W. Johnson, C. Dargemont, M. Seedorf and J. Warner for antibodies. We thank the GenePool at the University of Edinburgh for performing the MiSeq sequencing, and E. Petfalski for performing the initial crosslinking experiments. We thank E. Thomson for careful reading the manuscript. This work was supported by a postdoctoral fellowship from Alexander von Humboldt Foundation to Y.M., and by the Wellcome Trust to S.G. and D.T. (077248), and by grants from the Deutsche Forschungsgemeinschaft to E.H. (DFG Hu363/10-4).

**Author Contributions** Experiments were designed and the data were interpreted by Y.M. and E.H.; all experiments except CRAC analysis were performed by Y.M.; CRAC experiments and data analyses were performed by S.G. in collaboration with D.T.; M.T. constructed *rea1* mutants and performed the *in vitro* release assay of *rea1* mutants, the ctNug2 complementation assay and the immunodepletion assay; R.-G.M. developed the methods of the *in vitro* assay for nucleotide binding and GTPase activity measurement; the manuscript was written by Y.M. and E.H.; all authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.H. (ed.hurt@bzh.uni.heidelberg.de).



## METHODS

**Yeast strains and genetic methods.** The *S. cerevisiae* strains used in this study are listed in Extended Data Table 1. Gene disruption and C-terminal tagging were performed as previously described<sup>29,30</sup>.

**Plasmid constructs.** All recombinant DNA techniques were performed according to standard procedures using *E. coli* DH5 $\alpha$  for cloning and plasmid propagation. Site-directed mutagenesis was performed by overlap-extension PCR. All cloned DNA fragments generated by PCR amplification were verified by sequencing. Plasmids used in this study are listed in Extended Data Table 2.

**CRAC analysis.** The CRAC experiments were performed as described<sup>12</sup> using the Nug2- and Nmd3-HTP (His<sub>6</sub>-TEV-ProtA) strain. CRAC data were processed using pyCRAC (S. Webb, R. D. Hector, G. Kudla and S.G., manuscript submitted). Cells were ultraviolet-irradiated in the Megatron UV chamber<sup>1</sup> at a dose of 1.6 J cm<sup>-2</sup> and processed as described<sup>12,31</sup>. The cDNAs from the Nug2 CRAC data were cloned into pCR4-TOPO (Invitrogen), and inserts were sequenced by Sanger sequencing. The cDNAs originating from Nmd3 CRAC experiments were sequenced on the Illumina MiSeq system (single-end 50b), according to manufacturer's procedures. The MiSeq CRAC data were processed using the pyCRAC software suite (S. Webb, R. D. Hector, G. Kudla and S.G., manuscript submitted; <https://bitbucket.org/sgrann/pycrac>). To remove potential PCR duplicates, the Nmd3 MiSeq data was collapsed using pyFastqDuplicateRemover. Reads subsequently mapped to the yeast genomic reference sequence (version 2008) using novoalign (<http://www.novocraft.com>). Plots of reads aligned to the 35S reference sequence were generated using pyPileup and GNUplot. Adapters using this experiment are listed in Extended Data Table 3.

**Expression and purification of ctNug2.** The gene encoding *C. thermophilum* Nug2 (UniProtKB/TrEMBL accession: G0SXB1\_CHATD) was cloned from cDNA by standard procedures as recently described<sup>19</sup>. Subsequently, the ctNug2 was inserted into yeast or *E. coli* expression plasmids (see below). Because the C-terminal extension of ctNug2 (511–627 amino acids) is not conserved (Extended Data Fig. 1), ctNug2 from 1 to 510 amino acids was cloned into pET21 vector for the *in vitro* experiments. ctNug2 was expressed by using pET-ctNug2-510-His<sub>6</sub> plasmid in *E. coli* Rosetta-DE3 cells. Transformed cells were grown at 23 °C in LB medium until they reached an absorbance at 600 nm ( $A_{600\text{ nm}}$ ) of 0.6, isopropyl- $\beta$ -D-thiogalactoside (IPTG) was added to a final concentration of 0.1 mM. The cells were grown for an additional 3 h and then collected by centrifugation and stored frozen at –80 °C. Frozen pellets were resuspended in buffer KCl<sub>200</sub> (50 mM Tris, pH 8.0, 200 mM KCl, 5% glycerol, 0.01% NP-40 and 2 mM  $\beta$ -mercaptoethanol) with protease-inhibitor cocktail, and were broken by sonication (BANDELIN sonopuls 3200 with TITANTELLER TT13) on ice. Sonication was performed under these conditions: amplitude: 50%, 3 s on, 8 s off, processed for 10 min. The lysate was centrifuged at 39,000g for 30 min at 4 °C. The supernatant fraction was applied to a SP-sepharose column, and washed with buffer KCl<sub>200</sub>. ctNug2-His<sub>6</sub> was eluted by buffer KCl<sub>200</sub> containing 300 mM KCl. Next, the eluate fraction was applied to a Ni<sup>2+</sup>-NTA column, and the column was washed with buffer KCl<sub>200</sub>. ctNug2-His<sub>6</sub> was eluted with buffer KCl<sub>200</sub> containing 250 mM imidazole, before it was finally dialysed against buffer KCl<sub>200</sub>.

**Measurement of GTPase activity by single-turnover reactions.** The GTPase activity experiments were performed as previously described<sup>32</sup>. ctNug2, ctNug2K339R or ctNug2G380A (1  $\mu$ M) were incubated with a final concentration of 0.1  $\mu$ M GTP containing 750 nCi of [ $\gamma$ -<sup>32</sup>P]-labelled GTP in buffer KCl<sub>300</sub> (50 mM Tris, pH 8.0, 300 mM KCl, 10 mM MgCl<sub>2</sub> and 1 mM dithiothreitol (DTT)) or in buffer NaCl<sub>300</sub> (50 mM Tris, pH 8.0, 300 mM NaCl, 10 mM MgCl<sub>2</sub> and 1 mM DTT) for the indicated time at 30 °C. After the reaction, the hydrolysed  $\gamma$ -phosphate was separated by thin-layer chromatography.

**Guanine-nucleotide-binding experiments.** ctNug2, ctNug2K339R or ctNug2G380A (1  $\mu$ M) were incubated with 0.1  $\mu$ M MANT-GTP or MANT-GDP in buffer KCl<sub>300</sub> (50 mM Tris, pH 8.0, 300 mM KCl, 60 mM MgCl<sub>2</sub>, 20 mM EDTA and 1 mM DTT). MANT-GTP or MANT-GDP are analogues of natural GTP or GDP, where either the ribose 2'-hydroxy or the 3'-hydroxy group has been esterified by the fluorescent methylisatoic acid with an excitation/emission = 355/448 nm. The fluorescence quantum yield of MANT fluorophore is very low in water and increases significantly in non-polar solvents or after binding to most proteins. This highly environmental sensitive fluorescence of MANT makes MANT-GTP/GDP useful for directly detecting the nucleotide-protein interactions. Accordingly, it was excited at 355 nm with a xenon lamp, and emission spectra were recorded between 385 and 600 nm with a 5-nm increment step using a Synergy 4 spectrophotometer (BioTek).

**In vitro release assay.** The Rix1 particle was affinity purified using IgG beads via Rix1-TAP from yeast strain (Rix1-TAP, Rpl3-Flag) expressing NUG2, nug2(K328R)

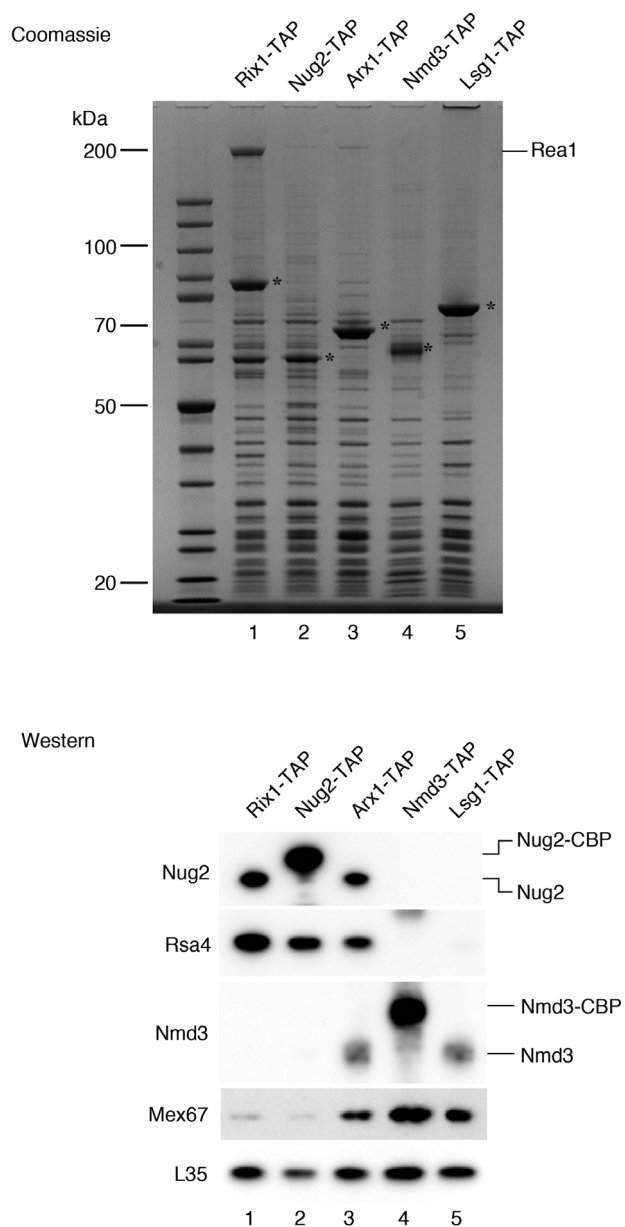
and nug2(G369A) followed by TEV protease cleavage at 4 °C to release the Rix1 particle. The TEV eluate (that is, the released Rix1 particle) was incubated with 4 mM ATP or 4 mM GTP at 23 °C for 1 h. After ATP or GTP treatment, the 60S particle was re-purified via affinity purification of Rpl3-Flag using Flag beads. Buffer KCl<sub>100</sub> (50 mM Tris, pH 8.0, 100 mM KCl, 10 mM MgCl<sub>2</sub> and 1 mM DTT) or buffer NaCl<sub>100</sub> (50 mM Tris, pH 8.0, 100 mM NaCl, 10 mM MgCl<sub>2</sub> and 1 mM DTT) were used. Affinity purifications were performed as previously described<sup>33</sup>.

**Immunodepletion of Rix1 by Flag immunoprecipitation.** The Nug2 particle was affinity purified from yeast strain (Nug2-TAP, Rix1-Flag) via IgG beads. The TEV eluate was incubated twice with Flag beads at 4 °C for 30 min each to deplete the Rix1-associated Nug2 particle. The flow-through was used for the final calmodulin purification step.

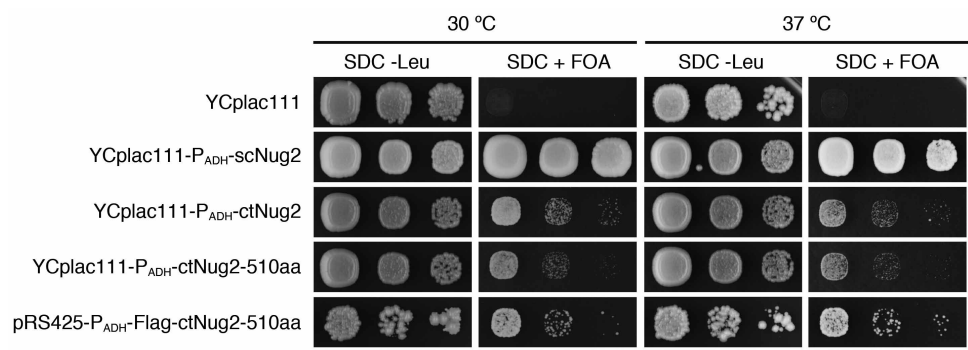
**Miscellaneous.** Further methods used in this study and previously described were TAP purifications of pre-60S particles<sup>34</sup>, sucrose gradient analysis to obtain ribosomal and polysomal profiles<sup>35</sup>, ribosomal export assays using the large subunit reporter Rpl25-eGFP monitored by fluorescence microscopy<sup>36</sup> and yeast three-hybrid analysis<sup>37</sup>. Antibodies used for western analysis in the following dilutions were anti-Nug2 (ref. 38) 1:10,000; anti-Rsa4 (ref. 39) 1:10,000; anti-Nmd3 (ref. 15) 1:5,000; anti-Mex67/Mtr2 (ref. 40) 1:10,000; anti-Rpl35 (ref. 41) 1:35,000; anti-Rpl3 (ref. 42) 1:5,000; anti-Nsa2 (ref. 43) 1:10,000; anti-Rpl10 (ref. 44) 1:1,000; anti-Nog1 (ref. 38) 1:30,000; anti-Rei1 (ref. 45) 1:5,000; goat-anti-mouse 1:3,000 (170-6516) and mouse-anti-rabbit horseradish-peroxidase-conjugated antibodies 1:3,000 (170-6515, both BioRad). Page ruler unstained protein ladder (Thermo Scientific) was used as a protein marker; Brilliant Blue G-Colloidal Concentrate Electrophoresis Reagent (Sigma-Aldrich) was used for Coomassie stain, and 4–12% NuPAGE Bis-Tris gels (Novex) together with NuPAGE MOPS SDS running buffer (Invitrogen) were used for SDS-PAGE. Wild-type yeast strain W303 (ref. 46) and the yeast three-hybrid vector p3A-MS2-1 (ref. 47) were used.

29. Janke, C. *et al.* A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast* **21**, 947–962 (2004).
30. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).
31. Granneman, S., Petfalski, E. & Tollervey, D. A cluster of ribosome synthesis factors regulate pre-rRNA folding and 5.8S rRNA maturation by the Rat1 exonuclease. *EMBO J.* **30**, 4006–4019 (2011).
32. Ferreira-Cerca, S. *et al.* ATPase-dependent role of the atypical kinase Rio2 on the evolving pre-40S ribosomal subunit. *Nature Struct. Mol. Biol.* **19**, 1316–1323 (2012).
33. Bradatsch, B. *et al.* Structure of the pre-60S ribosomal subunit with nuclear export factor Arx1 bound at the exit tunnel. *Nature Struct. Mol. Biol.* **19**, 1234–1241 (2012).
34. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
35. Bařler, J. *et al.* Identification of a 60S preribosomal particle that is closely linked to nuclear export. *Mol. Cell* **8**, 517–529 (2001).
36. Hurt, E. *et al.* A novel in vivo assay reveals inhibition of ribosomal nuclear export in ran-cycle and nucleoporin mutants. *J. Cell Biol.* **144**, 389–401 (1999).
37. SenGupta, D. J. *et al.* A three-hybrid system to detect RNA-protein interactions in vivo. *Proc. Natl Acad. Sci. USA* **93**, 8496–8501 (1996).
38. Saveanu, C. *et al.* Sequential protein association with nascent 60S ribosomal particles. *Mol. Cell Biol.* **23**, 4449–4460 (2003).
39. de la Cruz, J., Sanz-Martinez, E. & Remacha, M. The essential WD-repeat protein Rsa4p is required for rRNA processing and intra-nuclear transport of 60S ribosomal subunits. *Nucleic Acids Res.* **33**, 5728–5739 (2005).
40. Gwizdek, C. *et al.* Ubiquitin-associated domain of Mex67 synchronizes recruitment of the mRNA export machinery with transcription. *Proc. Natl Acad. Sci. USA* **103**, 16376–16381 (2006).
41. Frey, S., Pool, M. & Seedorf, M. Scp160p, an RNA-binding, polysome-associated protein, localizes to the endoplasmic reticulum of *Saccharomyces cerevisiae* in a microtubule-dependent manner. *J. Biol. Chem.* **276**, 15905–15912 (2001).
42. Vilardell, J. & Warner, J. R. Ribosomal protein L32 of *Saccharomyces cerevisiae* influences both the splicing of its own transcript and the processing of rRNA. *Mol. Cell Biol.* **17**, 1959–1965 (1997).
43. Lebreton, A., Saveanu, C., Decourty, L., Jacquier, A. & Fromont-Racine, M. Nsa2 is an unstable, conserved factor required for the maturation of 27 SB pre-rRNAs. *J. Biol. Chem.* **281**, 27099–27108 (2006).
44. Bussiere, C., Hashem, Y., Arora, S., Frank, J. & Johnson, A. W. Integrity of the P-site is probed during maturation of the 60S ribosomal subunit. *J. Cell Biol.* **197**, 747–759 (2012).
45. Lebreton, A. *et al.* A functional network involved in the recycling of nucleocytoplasmic pre-60S factors. *J. Cell Biol.* **173**, 349–360 (2006).
46. Thomas, B. J. & Rothstein, R. Elevated recombination rates in transcriptionally active DNA. *Cell* **56**, 619–630 (1989).
47. Bařler, J., Kallas, M. & Hurt, E. The NUG1 GTPase reveals and N-terminal RNA-binding domain that is essential for association with 60 S pre-ribosomal particles. *J. Biol. Chem.* **281**, 24737–24744 (2006).





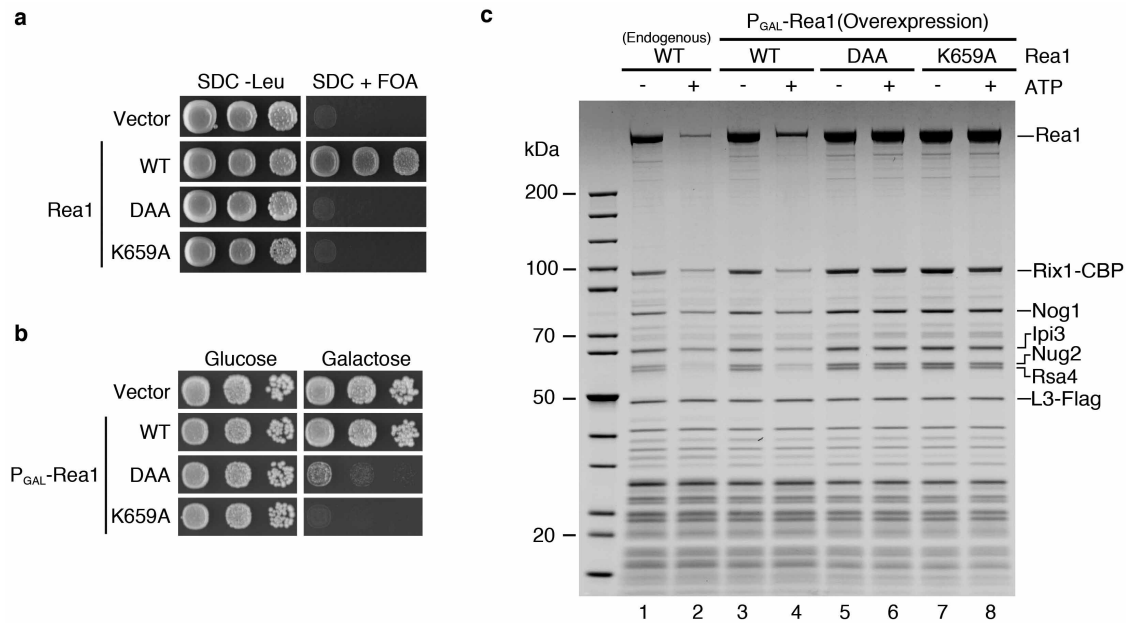
**Extended Data Figure 2 | Nug2 and Nmd3 are not found on the same pre-60S particles.** Indicated different TAP-tagged bait proteins were affinity purified from yeast wild-type cells. The final eluates were analysed by SDS-PAGE and Coomassie staining (top), and by western blotting using the indicated antibody (bottom). Asterisks mark the position of each bait protein. Rea1 has been identified by mass spectrometry. All affinity purifications and western analyses were performed at least twice, yielding highly reproducible data sets.



**Extended Data Figure 3 | ctNug2 can complement the lethal phenotype of a *nug2Δ* null mutant.** Serial dilutions of the yeast Nug2 shuffle strain (*MATa*, *ade2*, *ade3*, *his3*, *ura3*, *leu2*, *trp1*, *nug2::kanMX4*, pHT4467-NUG2) transformed with empty plasmid, yeast *ScNUG2*, *ctNUG2* or *ctNUG2-510* (truncation of the non-conserved C-terminal extension; see Extended Data

Fig. 1) under the control of the constitutive *ADH1* promoter in a single-copy-number (YCplac111) or multi-copy-number (pRS425) plasmid (see Supplementary Table 2) were spotted on SDC–Leu (loading control) and SDC plates containing 5-FOA at indicated temperatures for 6 days. Note that *ctNug2* only partially complements the *nug2* null mutant.

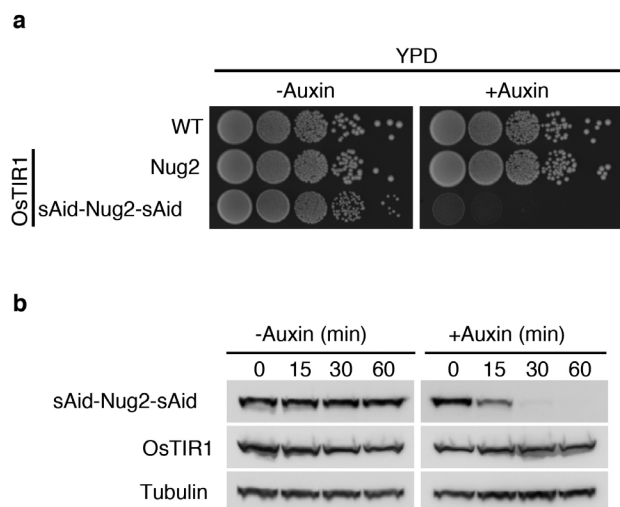




**Extended Data Figure 4 | Mutations in ATP-binding or MIDAS domain of Rea1 inhibit the release of Rsa4 and Nug2 from the pre-60S particle.**

**a, b,** Wild-type *REA1* and the *rea1* mutants mapping in the ATP-binding site of the AAA2 domain (Lys659Ala) or in the MIDAS domain (DAA)<sup>21</sup> were N-terminally tagged with eGFP and expressed in a *REA1* shuffle strain (**a**) or overexpressed under the control of the inducible *GAL1-10* promoter in *REA1* wild-type strain DS1-2b (**b**). Transformants were spotted in tenfold serial dilution steps on the indicated plates and incubated at 30 °C for 3 days. Both of the *rea1* mutant alleles do not complement the *rea1* null strain (**a**, SDC + 5-FOA) and cause a dominant-negative phenotype after overexpression by replacing endogenous Rea1 (**b**, galactose). **c**, Overnight pre-cultures were grown in SRC-Leu to prevent plasmid loss, followed by shifting cells

( $A_{600\text{ nm}} = 0.75$ ) to galactose medium (YPG) for 7 h. Rix1 particles, which were affinity purified from a Rix1-TAP, Rpl3-Flag strain containing either endogenous wild-type or overexpressed wild-type eGFP-Rea1, eGFP-Rea1(DAA) and eGFP-Rea1(Lys659Ala), were incubated with or without 4 mM ATP in KCl buffer, before the different *in vitro* matured pre-60S particles were re-isolated by affinity-purification via the Rpl3-Flag on Flag beads. Subsequently, the *in vitro* matured pre-60S particles (eluates) were analysed by SDS-PAGE and Coomassie staining. Relevant bands are indicated on the right. Note that in the case of the *rea1* mutants, the release of Nug2, Rsa4 but also Rea1 and the Rix1 complex is significantly inhibited. All *in vitro* assays were performed at least twice, yielding highly reproducible data sets.



**Extended Data Figure 5 | Nug2 depletion assay using the auxin-inducible degron system.** **a**, Growth of Nug2 auxin degron strains (sAid-Nug2-sAid) in the  $P_{ADH}$ -*OsTIR1* background on YPD plates with or without 500  $\mu$ M auxin (IAA). The cell growth of sAid-Nug2-sAid strain was inhibited by the addition of auxin. **b**, Western blotting of sAid-Nug2-sAid after auxin treatment. The depletion of sAid-Nug2-sAid occurred within about 30 min of auxin addition.

Extended Data Table 1 | Yeast strains used in this study

Name	Genotype	Source
W303	<i>Mata, ade2-1, his3-11,15, leu2-3,112, trp1-1, ura3-1, can1-100</i>	ref. <sup>46</sup>
DS1-2b	<i>MATa, his3-Δ200, leu2-Δ1, trp1-Δ63, ura3-52</i>	ref. <sup>21</sup>
Nug2-HTpA	W303, <i>Mata, NUG2-HTpA::His3MX6</i>	This work
Nmd3-HTpA	W303, <i>Mata, NMD3-HTpA::His3MX6</i>	This work
L40 coat	<i>Mata, his3-200, ura3-52, leu2-3, 112, trp1-1, ade2, LYS2:::(lexAop)-HIS3, LexA-MS2 coat (TRP1)</i>	ref. <sup>37</sup>
Nug2 Shuffle strain	<i>Mata, his3, ura3, leu2, trp1, lys2, nug2::kanMX4, pRS416-NUG2</i>	ref. <sup>35</sup>
Nug2 Shuffle strain ( for ctNug2 complementation)	<i>Mata, ade2, ade3, his3, ura3, leu2, trp1, nug2::kanMX4, pHT4467-NUG2</i>	ref. <sup>35</sup>
Nmd3 Shuffle strain	<i>Mata, his3, trp1, leu2, ura3, LYS2, ADE2, ADE3, NMD3::His3MX6, pRS316 (URA3) Nmd3,</i>	ref. <sup>14</sup>
Nug2 Shuffle strain, Ssf1-TAP	<i>Mata, his3, ura3, leu2, trp1, lys2, nug2::kanMX4, SSF1-TAP::TRP1, pRS416-NUG2</i>	This work
Nug2 Shuffle strain, Nsa1-TAP	<i>Mata, his3, ura3, leu2, trp1, lys2, nug2::kanMX4, NSA1-TAP::TRP1, pRS416-NUG2</i>	This work
Nug2 Shuffle strain, Rix1-TAP	<i>Mata, his3, ura3, leu2, trp1, lys2, nug2::kanMX4, RIX1-TAP::TRP1, pRS416-NUG2</i>	This work
Nug2 Shuffle strain, Arx1-TAP	<i>Mata, his3, ura3, leu2, trp1, lys2, nug2::kanMX4, ARX1-TAP::TRP1, pRS416-NUG2</i>	This work
Nug2 Shuffle strain, Lsg1-TAP	<i>Mata, his3, ura3, leu2, trp1, lys2, nug2::kanMX4, LSG1-TAP::TRP1, pRS416-NUG2</i>	This work
Nug2 Shuffle strain, Rix1-TAP, L3-Flag	<i>Mata, his3, ura3, leu2, trp1, lys2, nug2::kanMX4, RIX1-TAP::TRP1, RPL3-FLAG::natNT2, pRS416-NUG2</i>	This work
Rix1-TAP, L3-Flag	DS1-2b, <i>Mata, RIX1-TAP::TRP1, RPL3-FLAG::natNT2</i>	This work
Rea1 shuffle strain	W303 <i>Mata rea1::kanMX6 YCG-YLR106c</i>	ref. <sup>21</sup>
Nug2 Rsa4 double shuffle strain	<i>Mata, his3, ura3, leu2, trp1, lys2, nug2::kanMX4, rsa4::His3MX4, pRS416-NUG2, pRS316-RSA4</i>	This work
Nug2 Rea1 double shuffle strain	<i>Mata, his3, ura3, leu2, trp1, lys2, nug2::kanMX4, rea1::kanMX4, pRS416-NUG2, pRS316-REA1</i>	This work

Extended Data Table 2 | Plasmids used in this study

Name	Features	Source
pACTII-NUG2	2 $\mu$ , <i>LEU2</i> , P <sub>ADHI</sub> , T <sub>ADHI</sub> , <i>G4AD-NUG2</i>	Ref. <sup>21</sup>
p3A-MS2-1	2 $\mu$ , <i>URA3</i> , <i>ADE2</i> , P <sub>PolIII</sub> , <i>MS2</i> sites, T <sub>PolIII</sub>	Ref. <sup>47</sup>
p3A-MS2-25SH25	2 $\mu$ , <i>URA3</i> , <i>ADE2</i> , P <sub>PolIII</sub> , <i>MS2-25SH25</i> sites, T <sub>PolIII</sub>	This study
p3A-MS2-25SH38	2 $\mu$ , <i>URA3</i> , <i>ADE2</i> , P <sub>PolIII</sub> , <i>MS2-25SH38</i> , T <sub>PolIII</sub>	This study
p3A-MS2-25SH69-71	2 $\mu$ , <i>URA3</i> , <i>ADE2</i> , P <sub>PolIII</sub> , <i>MS2-25SH69-71</i> , T <sub>PolIII</sub>	This study
p3A-MS2-25SH89	2 $\mu$ , <i>URA3</i> , <i>ADE2</i> , P <sub>PolIII</sub> , <i>MS2-25SH89</i> , T <sub>PolIII</sub>	This study
p3A-MS2-25SH90-92	2 $\mu$ , <i>URA3</i> , <i>ADE2</i> , P <sub>PolIII</sub> , <i>MS2-25SH90-92</i> , T <sub>PolIII</sub>	This study
pCM185-NUG2	<i>CEN</i> , <i>TRP1</i> , P <sub>tetO7</sub> - <i>NUG2</i>	This study
pCM185-nug2K328R	<i>CEN</i> , <i>TRP1</i> , P <sub>tetO7</sub> - <i>nug2K328R</i>	This study
pCM185-nug2G369A	<i>CEN</i> , <i>TRP1</i> , P <sub>tetO7</sub> - <i>nug2g369A</i>	This study
pCM190-NUG2	2 $\mu$ , <i>URA3</i> , P <sub>tetO7</sub> - <i>NUG2</i>	This study
pCM190-nug2K328R	2 $\mu$ , <i>URA3</i> , P <sub>tetO7</sub> - <i>nug2K328R</i>	This study
pCM190-nug2G369A	2 $\mu$ , <i>URA3</i> , P <sub>tetO7</sub> - <i>nug2G369A</i>	This study
pRS313-NUG2	<i>CEN</i> , <i>HIS3</i> , <i>NUG2</i>	This study
pRS313-nug2K328R	<i>CEN</i> , <i>HIS3</i> , <i>nug2K328R</i>	This study
pRS313-nug2G369A	<i>CEN</i> , <i>HIS3</i> , <i>nug2G369A</i>	This study
pRS315-NUG2-TAP	<i>CEN</i> , <i>LEU2</i> , <i>NUG2-TAP</i>	Ref. <sup>35</sup>
pRS315-nug2K328R-TAP	<i>CEN</i> , <i>LEU2</i> , <i>nug2K328R-TAP</i>	This study
pRS314-NMD3-TAP	<i>CEN</i> , <i>TRP1</i> , <i>NMD3-TAP</i>	This study
YCplac111-GFP-REA1	<i>CEN</i> , <i>LEU2</i> , <i>GFP-REA1</i>	Ref. <sup>21</sup> & This Study
YCplac111-GFP-rea1DAA	<i>CEN</i> , <i>LEU2</i> , <i>GFP-rea1DAA</i>	Ref. <sup>21</sup> & This Study
YCplac111-GFP-REA1K659A	<i>CEN</i> , <i>LEU2</i> , <i>GFP-rea1K659A</i>	This study
YCplac111G-GFP-REA1	<i>CEN</i> , <i>LEU2</i> , P <sub>GALI-10</sub> - <i>GFP-REA1</i>	Ref. <sup>21</sup> & This Study
YCplac111G-GFP-rea1DAA	<i>CEN</i> , <i>LEU2</i> , P <sub>GALI-10</sub> - <i>GFP-rea1DAA</i>	Ref. <sup>21</sup> & This Study
YCplac111G-GFP-REA1K659A	<i>CEN</i> , <i>LEU2</i> , P <sub>GALI-10</sub> - <i>GFP-rea1K659A</i>	This study
YCplac111-P <sub>ADH</sub> -NUG2	<i>CEN</i> , <i>LEU2</i> , P <sub>ADHI</sub> - <i>NUG2</i>	This study
YCplac111-P <sub>ADH</sub> -ctNUG2	<i>CEN</i> , <i>LEU2</i> , P <sub>ADHI</sub> - <i>ctNUG2</i>	This study
YCplac111-P <sub>ADH</sub> -ctNug2-510	<i>CEN</i> , <i>LEU2</i> , P <sub>ADHI</sub> - <i>ctNUG2-510</i> (1-510aa)	This study
pRS425-P <sub>ADH</sub> -Flag-ctNUG2-510	2 $\mu$ , <i>LEU2</i> , P <sub>ADHI</sub> - <i>ctNUG2-510</i> (1-510aa)	This study
pRS314-RFP- <i>NOP1</i> - <i>RPL25</i> -GFP	<i>CEN</i> , <i>TRP1</i> , <i>mRFP-NOP1</i> , <i>RPL25-eGFP</i>	Ref. <sup>21</sup>
pRS314-RFP- <i>NOP1</i> - <i>RPS3</i> -GFP	<i>CEN</i> , <i>TRP1</i> , <i>mRFP-NOP1</i> , <i>RPS3-eGFP</i>	Ref. <sup>21</sup>
pRS314-rsa4-1	<i>CEN</i> , <i>TRP1</i> , <i>rsa4-1</i>	Ref. <sup>21</sup>
YCplac22-rea1DTS	<i>CEN</i> , <i>TRP1</i> , <i>rea1-DTS</i>	Ref. <sup>21</sup> & This Study
pRS313-sAid-Nug2-sAid, P <sub>ADH</sub> -OsTIR1	<i>CEN</i> , <i>HIS3</i> , <i>sAid-NUG2-sAid</i> , P <sub>ADHI</sub> - <i>OsTIR1-myc</i> <sub>9</sub>	This study
pET-ctNug2-510-His <sub>6</sub>	<i>E. coli</i> expression vector, Amp, <i>ctNUG-His<sub>6</sub></i> (1-510aa)	This study
pET-ctNugK339R-510-His <sub>6</sub>	<i>E. coli</i> expression vector, Amp, <i>ctnug2K339R-His<sub>6</sub></i> (1-510aa)	This study
pET-ctNug2-G380A-510-His <sub>6</sub>	<i>E. coli</i> expression vector, Amp, <i>ctnug2G380A-His<sub>6</sub></i> (1-510aa)	This study



**Extended Data Table 3 | Adapters used for the CRAC experiments**

Name	Sequence
<u>5' linkers:</u>	
Nug2 CRAC	5'-invddT-GTTCArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC-OH-3'
Nmd3 CRAC 1	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrUrCrUrCrUrArGrC-OH-3'
Nmd3 CRAC 2	5'-invddT-ACACrGrArCrGrCrUrCrUrUrCrCrGrArUrCrUrNrNrNrCrArCrUrArGrC-OH-3'
<u>3' linker:</u>	
miRCat	5'-AppTGGAATTCTCGGGTGCCAAGddC-3'
<u>PCR oligos used:</u>	
P5_forward	5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'
Pairedendmircat	5'-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTG-
Reverse	GCCTTGGCACCCGAGAATTCC-3'.

5'-invddT denotes an inverted dideoxy thymidine; N denotes random nucleotide sequences.

# N<sup>6</sup>-methyladenosine-dependent regulation of messenger RNA stability

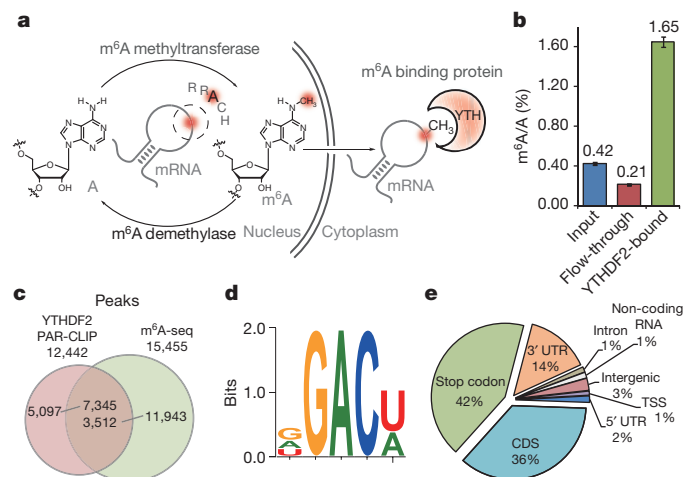
Xiao Wang<sup>1</sup>, Zhike Lu<sup>1</sup>, Adrian Gomez<sup>1</sup>, Gary C. Hon<sup>2</sup>, Yanan Yue<sup>1</sup>, Dali Han<sup>1</sup>, Ye Fu<sup>1</sup>, Marc Parisien<sup>3</sup>, Qing Dai<sup>1</sup>, Guifang Jia<sup>1,4</sup>, Bing Ren<sup>2</sup>, Tao Pan<sup>3</sup> & Chuan He<sup>1</sup>

N<sup>6</sup>-methyladenosine (m<sup>6</sup>A) is the most prevalent internal (non-cap) modification present in the messenger RNA of all higher eukaryotes<sup>1,2</sup>. Although essential to cell viability and development<sup>3–5</sup>, the exact role of m<sup>6</sup>A modification remains to be determined. The recent discovery of two m<sup>6</sup>A demethylases in mammalian cells highlighted the importance of m<sup>6</sup>A in basic biological functions and disease<sup>6–8</sup>. Here we show that m<sup>6</sup>A is selectively recognized by the human YTH domain family 2 (YTHDF2) 'reader' protein to regulate mRNA degradation. We identified over 3,000 cellular RNA targets of YTHDF2, most of which are mRNAs, but which also include non-coding RNAs, with a conserved core motif of G(m<sup>6</sup>A)C. We further establish the role of YTHDF2 in RNA metabolism, showing that binding of YTHDF2 results in the localization of bound mRNA from the translatable pool to mRNA decay sites, such as processing bodies<sup>9</sup>. The carboxy-terminal domain of YTHDF2 selectively binds to m<sup>6</sup>A-containing mRNA, whereas the amino-terminal domain is responsible for the localization of the YTHDF2–mRNA complex to cellular RNA decay sites. Our results indicate that the dynamic m<sup>6</sup>A modification is recognized by selectively binding proteins to affect the translation status and lifetime of mRNA.

Messenger RNA is central to the flow of genetic information. Regulatory elements (for example, AU-rich element, iron-responsive element), in the form of short sequence or structural motif imprinted in mRNA, are known to control the time and location of translation and degradation processes<sup>10</sup>. Reversible and dynamic methylation of mRNA could add another layer of more sophisticated regulation to the primary sequence<sup>2,11</sup>. m<sup>6</sup>A, a prevalent internal modification in the messenger RNA of all eukaryotes, is post-transcriptionally installed by m<sup>6</sup>A methyltransferase (for example, MT-A70, Fig. 1a) within the consensus sequence of G(m<sup>6</sup>A)C (70%) or A(m<sup>6</sup>A)C (30%)<sup>12</sup>. The loss of MT-A70 leads to apoptosis in human HeLa cells<sup>13</sup>, and significantly impairs development in *Arabidopsis*<sup>4</sup> and in *Drosophila*<sup>5</sup>. Our recent discoveries of m<sup>6</sup>A demethylases FTO (fat mass and obesity-associated protein)<sup>7</sup> and ALKBH5<sup>8</sup> demonstrate that this RNA methylation is reversible and may dynamically control mRNA metabolism. The recently revealed m<sup>6</sup>A transcriptomes (methylome) in human cells and mouse tissues showed m<sup>6</sup>A enrichments within long exons and around stop codons<sup>14,15</sup>, further suggesting fundamental regulatory roles of m<sup>6</sup>A. However, despite these progresses the exact function of m<sup>6</sup>A remains to be elucidated.

Whereas methyltransferase may serve as the 'writer' and demethylases (FTO and ALKBH5) act as the 'eraser' of m<sup>6</sup>A on mRNA, potential m<sup>6</sup>A-selective-binding proteins could represent the 'reader' of the m<sup>6</sup>A modification and exert regulatory functions through selective recognition of methylated RNA. Here, we show that the YTH-domain family member 2 (YTHDF2), initially found in pull-down experiments using m<sup>6</sup>A-containing RNA probes<sup>14</sup>, selectively binds m<sup>6</sup>A-methylated mRNA and controls RNA decay in a methylation-dependent manner.

The YTH domain family is widespread in eukaryotes and known to bind single-stranded RNA with the conserved YTH domain (>60% identity) located at the C terminus<sup>16,17</sup>. In addition to previously reported YTHDF2 and YTHDF3<sup>14</sup>, we also discovered YTHDF1 as another m<sup>6</sup>A-selective binding protein by using methylated RNA bait containing the known consensus sites of G(m<sup>6</sup>A)C and A(m<sup>6</sup>A)C versus unmethylated control (Extended Data Fig. 1a). Further, highly purified poly(A)-tailed RNAs were incubated with recombinant glutathione-S-transferase (GST)-tagged YTHDF1–3 and then separated by GST-affinity column. By using a previously reported liquid chromatography-tandem mass spectrometry (LC-MS/MS) method<sup>7,8</sup>, we found that the m<sup>6</sup>A-containing RNAs were greatly enriched in the YTHDF-bound portion and diminished in the flow-through portion (Fig. 1b and Extended Data Fig. 1b). Gel-shift assay revealed that YTHDF2 has a 16-fold higher binding affinity to methylated probe compared to the unmethylated one, as well as a slight preference to the consensus sequence (Extended Data Fig. 1c, d). This protein was selected for subsequent characterization because it has a high selectivity to m<sup>6</sup>A, and was thought to be associated with human longevity<sup>18</sup>.



**Figure 1** | YTHDF2 selectively binds m<sup>6</sup>A-containing RNA. **a**, Illustration of m<sup>6</sup>A methyltransferase, demethylase and binding proteins. RRACH is the extended m<sup>6</sup>A consensus motif, where R is G or A and H is not G. **b**, LC-MS/MS showing m<sup>6</sup>A enrichment in GST–YTHDF2-bound mRNA while depleted in the flow-through portion. Error bars, mean ± s.d., *n* = 2, technical replicates. **c**, Overlap of peaks identified through YTHDF2-based PAR-CLIP and the m<sup>6</sup>A-seq peaks in the same cell line. **d**, Binding motif identified by MEME with PAR-CLIP peaks (*P* = 3.0 × 10<sup>-46</sup>, 381 sites were found under this motif out of top 1,000 scored peaks). **e**, Pie chart depicting the region distribution of YTHDF2-binding sites identified by PAR-CLIP, TSS (200-bp window from the transcription starting site), stop codon (400-bp window centred on stop codon).

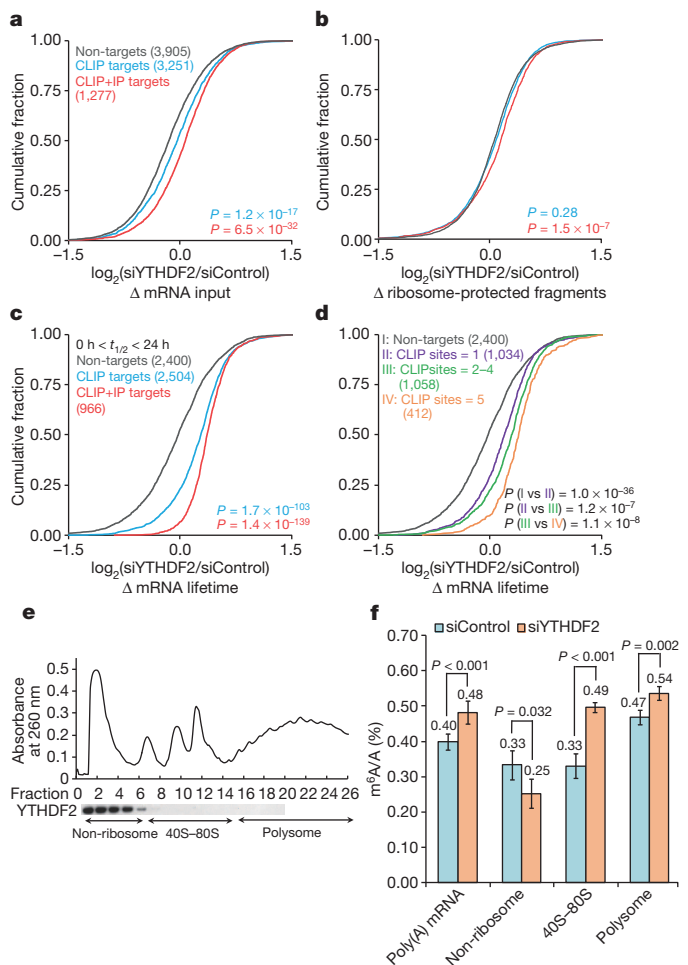
<sup>1</sup>Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA. <sup>2</sup>Ludwig Institute for Cancer Research, Department of Cellular and Molecular Medicine, UCSD Moores Cancer Center and Institute of Genome Medicine, University of California, San Diego School of Medicine, 9500 Gilman Drive, La Jolla, California 92093-0653, USA. <sup>3</sup>Department of Biochemistry and Molecular Biology and Institute for Biophysical Dynamics, The University of Chicago, 929 East 57th Street, Chicago, Illinois 60637, USA. <sup>4</sup>Department of Chemical Biology and Synthetic and Functional Biomolecules Center, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China.

We next applied two independent methods to identify RNAs that are the binding partners of YTHDF2: (1) photoactivatable ribonucleoside crosslinking and immunoprecipitation (PAR-CLIP)<sup>19</sup> to locate the binding sites of YTHDF2; (2) sequencing profiling of the RNA of immunopurified ribonucleoprotein complex (RNP) (RIP-seq)<sup>20</sup> to extract cellular YTHDF2–RNA complexes. Approximately 10,000 crosslinked clusters covering 3,251 genes were identified in PAR-CLIP (Extended Data Fig. 2a, b). Most are mRNA but 1% are non-coding RNA. Among 2,536 transcripts identified in RIP-seq, 50% overlap with PAR-CLIP targets (Extended Data Fig. 2b). We also performed m<sup>6</sup>A-seq for the poly(A)-tailed RNA from the same HeLa cell line and found that 59% (7,345 out of 12,442) of the PAR-CLIP peaks of YTHDF2 overlap with m<sup>6</sup>A peaks (Fig. 1c). As shown in Fig. 1d, the conserved motif revealed from the top 1,000 scored clusters matches the m<sup>6</sup>A consensus sequence of RRACH<sup>12,14</sup>, which strongly supports the binding of m<sup>6</sup>A by YTHDF2 inside cells (see more motifs in Extended Data Fig. 2c–e). Coinciding with the previously reported pattern of m<sup>6</sup>A peaks<sup>14,15</sup>, YTHDF2 PAR-CLIP peaks showed enrichment near the stop codon and in long exons (Extended Data Fig. 2f–h). YTHDF2 predominantly targets the stop codon region, the 3' untranslated region (3' UTR), and the coding region (CDS) (Fig. 1e), indicating that YTHDF2 may have a role in mRNA stability and/or translation.

To dissect the role of YTHDF2 we used ribosome profiling to assess the ribosome loading of each mRNA represented as ribosome-protected reads<sup>21,22</sup>. HeLa cells that were treated with YTHDF2 short interfering RNA (siRNA) (Extended Data Fig. 3a) as well as siRNA control were subsequently subjected to ribosome profiling with mRNA sequencing (mRNA-seq) performed on the same sample. Transcripts present (reads per kilobase per million reads (RPKM) > 1) in both ribosome profiling and mRNA-seq samples were analysed. These transcripts were then categorized as YTHDF2 PAR-CLIP targets (3,251), common targets of PAR-CLIP and RIP (1,277), and non-targets (3,905, absent from PAR-CLIP and RIP). A significant increase of input mRNA reads for YTHDF2 targets was observed in the YTHDF2 knockdown sample compared to the control ( $P < 0.001$ , Mann–Whitney  $U$ -test), without a noticeable change for non-targets (Fig. 2a). However, compared with the increase in mRNA level, the differences in the ribosome-protected fraction in the knockdown sample compared to the control were small (Fig. 2b). Thus, YTHDF2 knockdown led to apparently reduced translation efficiency of its targets as a result of accumulation of non-translating mRNA (Extended Data Fig. 3b), suggesting the primary role of YTHDF2 in RNA degradation.

Next, we performed RNA lifetime profiling by collecting and analysing RNA-seq data on YTHDF2 knockdown and control samples obtained at different time points after transcription inhibition with actinomycin D. Indeed, YTHDF2 knockdown led to prolonged (~30% in average) lifetimes of its mRNA targets in comparison with non-targets (Fig. 2c). Interestingly, we found that as the number of binding sites increase the stabilization of the RNA targets caused by YTHDF2 knockdown also increase significantly<sup>23</sup>: more than four sites have a larger extent of stabilization upon YTHDF2 knockdown than 2–4 sites, which have larger fold changes than targets with only one site (Fig. 2d and Extended Data Fig. 3c, Kruskal–Wallis test,  $P < 0.0001$ ); however, transcripts grouped according to binding region show similar fold-change indistinguishable in statistical test (Extended Data Fig. 3c, d).

Three pools of mRNAs exist in cytoplasm as defined by their engagement in translation<sup>24,25</sup> (Fig. 2e): non-ribosome mRNPs (mRNA–protein particles, with sedimentation coefficients of 20S–35S in sucrose gradient), translatable mRNA pool associated with ribosomal subunits (40S–80S), and actively translating polysome (>80S). YTHDF2 was observed to be present in non-ribosome fraction (Fig. 2e). After YTHDF2 knockdown, a 21% increase of the m<sup>6</sup>A/A ratio of the total mRNA was observed (Fig. 2f), confirming that the presence of YTHDF2 destabilizes the m<sup>6</sup>A-containing mRNA. YTHDF2 could affect localizing m<sup>6</sup>A-containing mRNA from a translatable pool to mRNPs. If so, the amount of methylated mRNA should decrease in mRNPs and



**Figure 2 | YTHDF2 destabilizes its cognate mRNAs.** **a–c**, Cumulative distribution of mRNA input (**a**), ribosome-protected fragments (**b**), and mRNA lifetime log<sub>2</sub> fold changes (**a, c**) between siYTHDF2 (YTHDF2 knockdown) and siControl (knockdown control) for non-targets (grey), PAR-CLIP targets (blue), and common targets of PAR-CLIP and RIP (red). **d**, The mRNA lifetime log<sub>2</sub> fold changes were further grouped and analysed on the basis of the number of CLIP sites on each transcript. The increased binding of YTHDF2 on its target transcript correlates with reduced mRNA lifetime.  $P$  values were calculated using two-sided Mann–Whitney or Kruskal–Wallis test (rank-sum test for the comparison of two or multiple samples, respectively). Detailed statistics are presented in Extended Data Fig. 3c. **e**, Western blotting of Flag-tagged YTHDF2 on each fraction of 10–50% sucrose gradient showing that YTHDF2 does not associate with ribosome. The fractions were grouped to non-ribosome mRNPs, 40S–80S, and polysome. **f**, Quantification of the m<sup>6</sup>A/A ratio of the total mRNA, non-ribosome portion, 40S–80S, and polysome by LC-MS/MS. Noticeable increases of the m<sup>6</sup>A/A ratio of the total mRNA, mRNA from 40S–80S, and mRNA from polysome were observed in the siYTHDF2 sample compared to control after 48 h. A reduced m<sup>6</sup>A/A ratio of mRNA isolated from the non-ribosome portion was observed in the same experiment.  $P$  values were determined using two-sided Student's  $t$ -test for paired samples. Error bars, mean  $\pm$  s.d., for poly(A)-tailed total mRNA input,  $n = 10$  (five biological replicates  $\times$  two technical replicates), and for the rest,  $n = 4$  (two biological replicates  $\times$  two technical replicates).

increase in the translatable pool upon YTHDF2 knockdown. Indeed, after YTHDF2 knockdown, the m<sup>6</sup>A/A ratio of mRNA isolated from mRNPs showed a 24% decrease and the ratio from the translatable pool demonstrated a 46% increase (Fig. 2f). We also observed a 14% increase of the m<sup>6</sup>A/A ratio of mRNA isolated from polysome after YTHDF2 knockdown (Fig. 2f), although it is worth noting that this model provided no prediction of the behaviour of polysome because the ribosome-loading number per transcript depends on the availability of both mRNA and free ribosomes. It should be also noted that the



observed  $m^6A/A$  ratio change does not seem to result from the protein level change of methyltransferase and demethylase as detected by western blotting (Extended Data Fig. 3e).

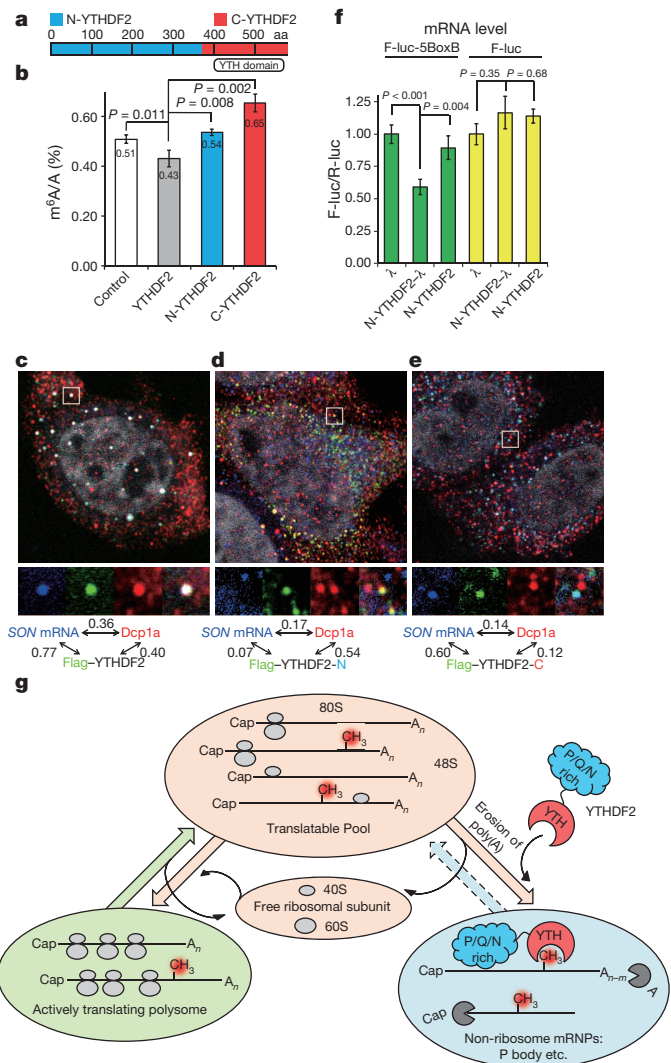
Three YTHDF2-targeted RNAs were selected for further validation: the *SON* mRNA has multiple CLIP peaks in CDS, the *CREBBP* mRNA has CLIP peaks at 3' UTR, and a non-coding RNA *PLAC2* (Extended Data Fig. 4a–d). As detected by gene-specific PCR with reverse transcription (RT-PCR), after 48 h YTHDF2 knockdown, all three RNA transcripts increased by more than 60% with prolonged lifetime; both *SON* and *CREBBP* showed redistribution from non-ribosome mRNP to translatable pool (Extended Data Fig. 4e–n). Furthermore, knockdown of the known  $m^6A$  methyltransferase *MT-A70* led to noticeably reduced binding of YTHDF2 to its targets and increased stability of the targets similar to that of the YTHDF2 knockdown (Extended Data Fig. 5).

To gain mechanistic understanding of the YTHDF2–mRNA interaction, we analysed the cellular distribution of YTHDF2 and found that YTHDF2 co-localizes with three markers (DCP1a, GW182 and DDX6) of processing bodies (P bodies) in the cytoplasm, where mRNA decay occurs (Extended Data Fig. 6a–j)<sup>26</sup>. YTHDF2 is composed of a C-terminal RNA-binding domain (C-YTHDF2) and a P/Q/N-rich N terminus (N-YTHDF2, Fig. 3a and Extended Data Fig. 6k)<sup>27,28</sup>. Whereas overexpression of YTHDF2 led to a reduced  $m^6A/A$  ratio of the total mRNA, overexpression of either N-YTHDF2 or C-YTHDF2 yielded an increased  $m^6A/A$  ratio (Fig. 3b), indicating that both domains are required for the YTHDF2-mediated mRNA decay. An *in vitro* pull-down experiment further showed that purified C-YTHDF2 is able to enrich  $m^6A$ -containing mRNA from total mRNA (Extended Data Fig. 6l). The spatial distribution of the *SON* mRNA relative to YTHDF2 and N- and C-YTHDF2 truncates were examined by fluorescence *in situ* hybridization (FISH) and fluorescence immunostaining in HeLa cells (Fig. 3c–e). The location of the *SON* mRNA showed a strong correlation with that of the full-length YTHDF2 (Fig. 3c) and C-YTHDF2 (Fig. 3e). In contrast, a much lower correlation was observed for the *SON* mRNA with N-YTHDF2 (Fig. 3d). In addition, the full-length YTHDF2 and N-YTHDF2 co-localized with DCP1a, but to a much lesser extent for C-YTHDF2, thereby indicating the role of N-YTHDF2 in P-body localization. Furthermore, the overexpression of C-YTHDF2 led to a reduced co-localization of the *SON* mRNA with DCP1a (Fig. 3e).

In further support of this mechanism, N-YTHDF2 was fused with  $\lambda$  peptide (N-YTHDF2- $\lambda$ ), which recognizes Box B RNA with a high affinity in a tether reporter assay<sup>29,30</sup>. Tethering N-YTHDF2- $\lambda$  to F-luc-5BoxB (five Box B sequence was inserted into the 3' UTR of the mRNA reporter) led to a significantly reduced mRNA level (Fig. 3f) and shortening (40%) of its lifetime compared with tethering controls of N-YTHDF2 or  $\lambda$  alone (Extended Data Fig. 7a–e). The reporter mRNA bound by N-YTHDF2- $\lambda$  possesses shorter poly(A) tail length in comparison with unbound portion, although a significant change of the deadenylation rate was not observed (Extended Data Fig. 7f–l). Together with the observation that YTHDF2 co-localizes with both deadenylation and decapping enzyme complexes (Extended Data Fig. 6), we propose a model (Fig. 3g) that consists of: (1) C-YTHDF2 selectively recognizes  $m^6A$ -containing mRNA less engaged with translation; (2) this binding of YTHDF2 to methylated mRNA happens in parallel or at a later stage of deadenylation; (3) N-YTHDF2 localizes the YTHDF2- $m^6A$ -mRNA complex to more specialized mRNA decay machineries (P bodies etc.) for committed degradation.

Functional clustering of YTHDF2 targets versus non-targets revealed that the main functions of YTHDF2-mediated RNA processing are gene expression (molecular function) as well as cell death and survival (cellular function, Extended Data Fig. 8a–d). After 72 h of YTHDF2 knockdown, the viability of HeLa cells reduced by 50% (Extended Data Fig. 8e, f), indicating that the YTHDF2-mediated RNA processing could have biological significance.

In summary, we present a transcriptome-wide identification of YTHDF2–RNA interaction and a mechanistic model for  $m^6A$  function



**Figure 3 | YTHDF2 affects *SON* mRNA localization in processing body (P-body).** **a**, Schematic of the domain architecture (aa, amino acids) of YTHDF2, N terminus of YTHDF2 (N-YTHDF2, aa 1–389, blue) and C terminus of YTHDF2 (C-YTHDF2, aa 390–end, red). **b**, Overexpression of full-length YTHDF2 led to reduced levels of  $m^6A$  after 24 h, whereas overexpression of N-YTHDF2 or C-YTHDF2 increased the  $m^6A/A$  ratio of the total mRNA. P values were determined using two-sided Student's *t*-test for paired samples. Error bars, mean  $\pm$  s.d.,  $n = 4$  (two biological replicates  $\times$  two technical replicates). **c–e**, Fluorescence *in situ* hybridization of *SON* mRNA and fluorescence immunostaining of DCP1a (P-body marker), Flag-tagged YTHDF2 (**c**), Flag-tagged C-YTHDF2 (**d**), and Flag-tagged N-YTHDF2 (**e**). Full-length YTHDF2 and C-YTHDF2 co-localize with *SON* mRNA (bearing  $m^6A$ ) and the full-length YTHDF2 significantly increases the P-body localization of *SON* mRNA compared to N-YTHDF2 and C-YTHDF2. The numbers shown above figures are Pearson correlation coefficients of each channel pair with the scale of the magnified region (white frame) set as  $2 \mu m \times 2 \mu m$ . **f**, Tethering N-YTHDF2- $\lambda$  to a mRNA reporter F-luc-5BoxB led to a  $\sim 40\%$  reduction of the reporter mRNA level compared to tethering N-YTHDF2 or  $\lambda$  alone (green) and controls without BoxB (F-luc, yellow). P values were determined using two-sided Student's *t*-test for paired samples. Error bars, mean  $\pm$  s.d.,  $n = 6$  (F-luc-5BoxB) or 3 (F-luc). **g**, A proposed model of  $m^6A$ -dependent mRNA degradation mediated through YTHDF2. The three states of mRNAs in cytoplasm are defined by their engagement with ribosome using the sedimentation coefficient range in sucrose gradient:  $>80S$  for actively translating polysome;  $40S$ – $80S$  for translatable pool;  $20S$ – $35S$  for non-ribosome mRNPs.

mediated by this  $m^6A$ -binding protein, as the first functional demonstration of a  $m^6A$  reader protein. We show that YTHDF2 alters the distribution of the cytoplasmic states of several thousand  $m^6A$ -containing



mRNA. This present work demonstrates that reversible m<sup>6</sup>A deposition could dynamically tune the stability and localization of the target RNAs through m<sup>6</sup>A 'readers'.

## METHODS SUMMARY

m<sup>6</sup>A profiling, PAR-CLIP and RIP experiments were conducted as previously reported<sup>14,19,20</sup>. For ribosome profiling, RPF was obtained by micrococcal nuclease digestion followed by sucrose gradient (10–50%) separation. Complementary DNA libraries of RPF and mRNA input were constructed as previously described<sup>22</sup>. In RNA lifetime profiling, actinomycin D (5 µg ml<sup>-1</sup>) was added to stop transcription, and samples at 0, 3 and 6 h decay were collected. ERCC RNA spike-in control (Ambion) was added to each sample before the isolation of mRNA and library construction to correct the decrease of the whole mRNA population during RNA decay. All of the cDNA libraries were sequenced by using HiSeq 2000 (Illumina, single end, 100 bp) and at least two replicates were performed for each experiment (Extended Data Table 1). The deep sequencing data were mapped to Human genome version hg19 without any gaps and allowed for at most two mismatches. The PAR-CLIP binding sites were identified through kernel density estimation of T to C conversions. For RIP, transcripts that have more than twofold enrichment were identified as targets. For ribosome profiling and mRNA lifetime profiling, the average of the log<sub>2</sub>(siYTHDF2/siControl) values generated from two biological replicates were analysed and comparisons of independent replicates were summarized in Extended Data Fig. 9.

**Online Content** Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 31 December 2012; accepted 3 October 2013.**

**Published online 27 November 2013; corrected online 1 January 2014 (see full-text HTML version for details).**

1. Tuck, M. T. The formation of internal 6-methyladenine residues in eucaryotic messenger RNA. *Int. J. Biochem.* **24**, 379–386 (1992).
2. Jia, G., Fu, Y. & He, C. Reversible RNA adenosine methylation in biological regulation. *Trends Genet.* **29**, 108–115 (2013).
3. Clancy, M. J., Shambaugh, M. E., Timpte, C. S. & Bokar, J. A. Induction of sporulation in *Saccharomyces cerevisiae* leads to the formation of N<sup>6</sup>-methyladenosine in mRNA: a potential mechanism for the activity of the *IME4* gene. *Nucleic Acids Res.* **30**, 4509–4518 (2002).
4. Zhong, S. *et al.* MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* **20**, 1278–1288 (2008).
5. Hongay, C. F. & Orr-Weaver, T. L. *Drosophila* Inducer of MEiosis 4 (IME4) is required for Notch signaling during oogenesis. *Proc. Natl Acad. Sci. USA* **108**, 14855–14860 (2011).
6. Frayling, T. M. *et al.* A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007).
7. Jia, G. *et al.* N<sup>6</sup>-methyladenosine in nuclear RNA is a major substrate of the obesity-associated *FTO*. *Nature Chem. Biol.* **7**, 885–887 (2011).
8. Zheng, G. *et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).
9. Sheth, U. & Parker, R. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* **300**, 805–808 (2003).
10. Dandekar, T. & Bengert, P. *RNA Motifs and Regulatory Elements* 2nd edn, 1–11, (Springer, 2002).
11. He, C. Grand challenge commentary: RNA epigenetics? *Nature Chem. Biol.* **6**, 863–865 (2010).
12. Wei, C. M. & Moss, B. Nucleotide sequences at the N<sup>6</sup>-methyladenosine sites of HeLa cell messenger ribonucleic acid. *Biochemistry* **16**, 1672–1676 (1977).
13. Bokar, J. A., Shambaugh, M. E., Polayes, D., Matera, A. G. & Rottman, F. M. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N<sup>6</sup>-adenosine)-methyltransferase. *RNA* **3**, 1233–1247 (1997).
14. Dominissini, D. *et al.* Topology of the human and mouse m<sup>6</sup>A RNA methylomes revealed by m<sup>6</sup>A-seq. *Nature* **485**, 201–206 (2012).
15. Meyer, K. D. *et al.* Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
16. Stoilov, P., Rafalska, I. & Stamm, S. YTH: a new domain in nuclear proteins. *Trends Biochem. Sci.* **27**, 495–497 (2002).
17. Zhang, Z. *et al.* The YTH domain is a novel RNA binding domain. *J. Biol. Chem.* **285**, 14701–14710 (2010).
18. Cardelli, M. *et al.* A polymorphism of the *YTHDF2* gene (1p35) located in an Alu-rich genomic domain is associated with human longevity. *J. Gerontol. A* **61**, 547–556 (2006).
19. Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129–141 (2010).
20. Peritz, T. *et al.* Immunoprecipitation of mRNA-protein complexes. *Nature Protocols* **1**, 577–580 (2006).
21. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
22. Bazzini, A. A., Lee, M. T. & Giraldez, A. J. Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. *Science* **336**, 233–237 (2012).
23. Mukherjee, N. *et al.* Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol. Cell* **43**, 327–339 (2011).
24. Huang, R., Brown, C. Y. & Morris, D. R. In *mRNA Formation and Function* (ed. Richter, J. D.) Ch. 16 (Academic Press, 1997).
25. Shenton, D. *et al.* Global translational responses to oxidative stress impact upon multiple levels of protein synthesis. *J. Biol. Chem.* **281**, 29011–29021 (2006).
26. Kedersha, N. & Anderson, P. Mammalian stress granules and processing bodies. *Methods Enzymol.* **431**, 61–81 (2007).
27. Reijns, M. A., Alexander, R. D., Spiller, M. P. & Beggs, J. D. A role for Q/N-rich aggregation-prone regions in P-body localization. *J. Cell Sci.* **121**, 2463–2472 (2008).
28. Kato, M. *et al.* Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753–767 (2012).
29. Gehring, N. H., Neu-Yilik, G., Schell, T., Hentze, M. W. & Kulozik, A. E. Y14 and hUp3b form an NMD-activating complex. *Mol. Cell* **11**, 939–949 (2003).
30. Behm-Ansmant, I. *et al.* mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev.* **20**, 1885–1898 (2006).

**Acknowledgements** This work is supported by National Institutes of Health GM071440 (C.H.) and EUREKA GM088599 (T.P. and C.H.). The Mass Spectrometry Facility of the University of Chicago is funded by National Science Foundation (CHE-1048528). We thank A. E. Kulozik, W. Filipowicz and J. A. Steitz for providing the sequence and plasmids of the tether reporter. We thank Dr. G. Zheng and W. Clark for help in polysome profiling. We also thank S. F. Reichard for editing the manuscript.

**Author Contributions** C.H. conceived the project. X.W. designed and performed most experiments. Z.L. and X.W. performed data analyses of high-throughput sequencing data. A.G. assisted with the experiments. Y.Y. and D.H. conducted the experimental and data analysis part of m<sup>6</sup>A profiling, respectively. Y.F. performed the RNA-affinity pull-down experiment of YTHDF1 and YTHDF3. M.P. and G.J. provided valuable discussions. G.C.H. and B.R. performed high throughput sequencing. Q.D. assisted in m<sup>6</sup>A synthesis. X.W. and C.H. interpreted the results and wrote the manuscript with input from T.P.

**Author Information** RNA sequencing data were deposited in the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under accession number GSE49339 and the processed results were presented as Supplementary Table 1. Processed files were deposited in the Gene Expression Omnibus under accession no. GSE46705. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.H. ([chuanhe@uchicago.edu](mailto:chuanhe@uchicago.edu)).

## METHODS

**Plasmid construction and protein expression.** Recombinant YTHDF1-3 were cloned from commercial cDNA clones (Open Biosystems) into vector pGEX-4T-1. The primers used for subcloning (from 5' to 3'; F stands for forward primer; R stands for reverse primer) are listed: GST-YTHDF1-F, CGATCGAATTCATG TCGGCCACAGCG; GST-YTHDF1-R, CCATACTCGAGTCATTGTTTGTTCGACTCTGCC; GST-YTHDF2-F, CGTACGGATCCATGTCAGATTCTACT TACCCAG; GST-YTHDF2-R, CGATGCTCGAGTCATTTCCCACGACCTTG ACG; GST-YTHDF3-F, CGTACGGATCCATGTCAGCCACTAGCGGTG; GST-YTHDF3-R, CGTAGCTCGAGTCATTGTTTGTTCATTTCTCTCCCTAC.

The resulting clones were transfected into the *Escherichia coli* strain BL21 and expression was induced at 16 °C with 1 mM IPTG for 20 h. The pellet collected from 2 litres of bacteria culture was then lysed in 30 ml PBS-L solution (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 150 mM NaCl, pH 7.2, 1 mM PMSF, 1 mM DTT, 1 mM EDTA, 0.1% (v/v) Triton X-100) and sonicated for 10 min. After removing cell debris by centrifuge at 17,000g for 30 min, the supernatant were loaded to a GST superflow cartridge (Qiagen, 5 ml) and gradiently eluted by using PBS-EW (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 150 mM NaCl, pH 7.2, 1 mM DTT, 1 mM EDTA) as buffer A and TNGT (50 mM Tris, pH 8.0, 150 mM NaCl, 50 mM red, GSH, 0.05% Triton X-100) as buffer B. The crude products were further purified by gel-filtration chromatography in GF buffer (10 mM Tris, pH 7.5, 200 mM NaCl, 3 mM DTT and 5% glycerol). The yield was around 1–2 mg per litre of bacterial culture.

Flag-tagged YTHDF2 was cloned into vector pcDNA 3.0 (BamHI, XhoI, forward primer, CGTACGGATCCATGGATTACAAGGACGACGATGACAA TGTCGGCCAGCAGCC; reverse primer, CGATGCTCGAGTCATTTCCCACG ACCTTGACG). Flag-tagged YTHDF2 N-terminal domain was made by mutating E384 (GAA) to a stop codon (TAA) with a Stratagene QuikChange II site-directed mutagenesis kit (pcDNA-Flag-Y2N, forward primer, CTGGATCTACTCTTCATAA CCCCACCCAGTGTG; reverse primer, CAACACTGG TGGGGCTATGAGG GGAGTAGATCCAG). Flag-tagged YTHDF2 C-terminal domain was made by cloning amino acids from E384 to the end into vector pcDNA 3.0 (BamHI, XhoI, forward primer, CGTACGGATCCATGGATTACAAGGACGACGATGACAA GGAACCCACCC AGTGTT; reverse primer, CGATGCTCGAGTCATTTCCC ACACCTTGACG). Plasmids with high purity for mammalian cell transfection were prepared with a Maxiprep kit (Qiagen).

Tether reporter: pmirGlo Dual luciferase expression vector (Promega) was used to construct the tether reporter which contains firefly luciferase (F-luc) as the primary reporter and *Renilla* luciferase (R-luc) acting as a control reporter for normalization. F-luc-5BoxB mRNA reporter was obtained by inserting five Box B sequence (5BoxB) into the 3' UTR of F-luc (SacI and XhoI, the resulting plasmid was named as pmirGlo-5BoxB). The 5BoxB sequence<sup>29</sup> (see below) was PCR-amplified from PRL-5BoxB plasmid, which was provided by W. Filipowi (forward primer, CGATACGAGCTCTCCCTAAGTCCAACTACCAAAC; reverse primer, CTATGGCTCGAGATAATATCTCTCGATAGGGCCC; sequencing primer, GACGAGGTGCCTAAAGA)<sup>31</sup>.

The 5BoxB sequence: TTCCTAAGTCCAACTACTAACTGGGGATTCCCT GGGCCCTGAAGAAGGGCCCTCGACTAAGTCCAACTACTAACTGGGC CTTAAGAAGGGCCCATATAGGGCCCTGAAGAAGGGCCCTATCGAGG ACATATATCGAGTCAAGTCAACTAACTGGGGCCCTGAAGAAGGG CCCATATAGGGCCCTGAAGAAGGGCCCTATCGAGGATATTATCTCGAG.

To study the decay kinetics of F-luc-5BoxB, another reporter plasmid (pmirGlo-Ptight-5BoxB) was constructed by replacing the original human phosphoglycerate kinase promoter of F-luc with Ptight promoter (restriction sites: ApaI and BglII). Ptight promoter was PCR amplified from pTRE-Tight vector (Clontech; forward primer, CGTACAGATCTCGAGTTTACTCCCTATCAGT; reverse primer, CTG TAGGGCCCT TCTAATGTTTTGGCATCTTCCATCTCCAGGCGATCTG ACG; sequencing primer, AGCGGTGCGTACAATTAAGG). The resulting plasmid (pmirGlo-Ptight) was subjected to a second round of subcloning by inserting 5BoxB into the 3' UTR of F-luc (restriction sites: XbaI and SbfI) to generate pmirGlo-Ptight-5BoxB (forward primer, CGTACTCTAGATTCCCTAAGTCC AACTACCAAAC; reverse primer, CTATGGCCCTCAGGATAATATCTCTG ATAGGGCCC; sequencing primer, GACGAGGTGCCTAA AGA).

Tether effector:  $\lambda$  peptide sequence (MDAQTRRRERRAEKQAWKAAAN) was fused to the C terminus of N-YTHDF2 by subcloning N-YTHDF2 to pcDNA 3.0 with forward primer containing Flag-tag sequence and reverse primer containing  $\lambda$  peptide sequence (pcDNA-Flag-Y2N $\lambda$ , BamHI, XhoI; forward primer, GATACGG ATCCATGGATTACAAGGACGACGATGACAAGATGTCGGCCAGCAGCC; reverse primer, TATGGCTCGAGTCAGTTGCACTTTCCATTGAGCTTGT TTCTACGCGCAGCTCAGCTCGTCTGTTTGTGCTCATCACTGAGG GGAGTAGATCCAGAACC). The  $\lambda$  peptide control was designed with a Flag tag at N-terminal and a GGS spacer (pcDNA-Flag- $\lambda$ ). The primer pair that contains Flag-tagged  $\lambda$  peptide and sticky restriction enzyme sites (*Bam*HI, *Xho*I) was annealed and directly ligated to digested pcDNA 3.0 (forward primer, GAT

CCATGGATTACAAGGACGACGATGACAAGGGTGGTAGCATGGACGCA CAAACACGACGACGTGAGCGTCGCGCTGAGAAACAAGCTCAATGGAA AGCTGCAAACTAAC; reverse primer, GAGTTAGTTTGCAGCTTTCCATTG AGCTTGTTCCTCAGCGCAGCGCTCAGTCGTCGTGTTTGTGCGTCCATG CTACCACCCTTGTCATCGTCGTCCTTGTATCCATG).

**EMSA (electrophoretic mobility shift assay/gel shift assay).** The RNA probe was synthesized by a previously reported method with the sequence of 5'-AUGGGC CGUUAUCUGCUAAAAGGCUGCUUUUGGGGCUUGU-3' (X = A or m<sup>6</sup>A). After the synthesis, the RNA probe was labelled in a reaction mixture of 2  $\mu$ l RNA probe (1  $\mu$ M), 5  $\mu$ l 5  $\times$  T4 PNK buffer A (Fermentas), 1  $\mu$ l T4 PNK (Fermentas), 1  $\mu$ l [<sup>32</sup>P]ATP and 41  $\mu$ l RNase-free water (final RNA concentration 40 nM) at 37 °C for 1 h. The mixture was then purified by RNase-free micro bio-spin columns with bio-gel P30 in Tris buffer (Bio-Rad 732-6250) to remove hot ATP and other small molecules. To the elute, 2.5  $\mu$ l 20  $\times$  SSC (Promega) buffer was added. The mixture was heated to 65 °C for 10 min to denature the RNA probe, and then slowly cooled down to room temperature. GST-YTHDF1, GST-YTHDF2 and GST-YTHDF3 were diluted to concentration series of 200 nmol, 1  $\mu$ M, 5  $\mu$ M, 20  $\mu$ M and 100  $\mu$ M (or other indicated concentrations) in binding buffer (10 mM HEPES, pH 8.0, 50 mM KCl, 1 mM EDTA, 0.05% Triton-X-100, 5% glycerol, 10  $\mu$ g ml<sup>-1</sup> salmon DNA, 1 mM DTT and 40 U ml<sup>-1</sup> RNasin). Before loading to each well, 1  $\mu$ l RNA probe (4 nM final concentration) and 1  $\mu$ l protein (20 nM, 100 nM, 500 nM, 2  $\mu$ M or 10  $\mu$ M final concentration) were added and the solution was incubated on ice for 30 min. The entire 10  $\mu$ l RNA-protein mixture was loaded to the gel (Novex 4–20% TBE gel) and run at 4 °C for 90 min at 90 V. Quantification of each band was carried out by using a storage phosphor screen (K-Screen; Fuji film) and Bio-Rad Molecular Imager FX in combination with Quantity One software (Bio-Rad). The  $K_d$  (dissociation constant) was calculated with nonlinear curve fitting (Function Hyperbl) of Origin 8 software with  $y = P_1 \times x/(P_2 + x)$ , where  $y$  is the ratio of [RNA-protein]/[free RNA] + [RNA-protein],  $x$  is the concentration of the protein,  $P_1$  is set to 1 and  $P_2$  is  $K_d$ .

**Mammalian cell culture, siRNA knockdown and plasmid transfection.** Human HeLa cell line used in this study was purchased from ATCC (CCL-2) and grown in DMEM (Gibco, 11965) media supplemented with 10% FBS and 1% 100  $\times$  Pen Strep (Gibco). HeLa Tet-off cell line was purchased from Clontech and grown in DMEM (Gibco) media supplemented with 10% FBS (Tet system approved, Clontech), 1% 100  $\times$  Pen Strep (Gibco) and 200  $\mu$ g ml<sup>-1</sup> G418 (Clontech). AllStars negative control siRNA from Qiagen (1027281) was used as control siRNA in knockdown experiments. YTHDF2 siRNA was ordered from Qiagen as custom synthesis which targets 5'-AAGGACGTTCCCAATAGCCAA-3' near the N terminus of CDS. MT-A70 siRNA was ordered from Qiagen: 5'-CGTCAGTATCTTGGGCAAGTT-3'. Transfection was achieved by using Lipofectamine RNAiMAX (Invitrogen) for siRNA, and Lipofectamine 2000 for single type of plasmid or Lipofectamine LTX Plus (Invitrogen) for co-transfection of two or multiple types of plasmids (tethering assay) following the manufacturer's protocols.

**RNA isolation.** mRNA isolation for LC-MS/MS: total RNA was isolated from wild-type or transiently transfected cells with TRIzol reagent (Invitrogen). mRNA was extracted using PolyATtract mRNA Isolation Systems IV (Promega) followed by further removal of contaminated rRNA by using RiboMinus Transcriptome Isolation Kit (Invitrogen). mRNA concentration was measured by NanoDrop. Total RNA isolation for RT-PCR: following the instruction of RNeasy kit (Qiagen) in addition to DNase I digestion step. Ethanol precipitation: to the RNA solution being purified or concentrated, 1/10 volume of 3 M NaOAc, pH 5.5, 1  $\mu$ l glycogen (10 mg ml<sup>-1</sup>) and 2.7 volume of 100% ethanol were added, stored at -80 °C for 1 h to overnight, and then centrifuged at 15,000g for 15 min. After the supernatant was removed, the pellet was washed twice by using 1 ml 75% ethanol, and dissolved in the appropriate amount of RNase-free water as indicated.

**In vitro pull down.** 0.8  $\mu$ g mRNA (save 0.2  $\mu$ g from the same sample as input) and YTHDF1, YTHDF2, YTHDF3 or C-YTHDF2 (final concentration 500 nM) were diluted into 200  $\mu$ l IPP buffer (150 mM NaCl, 0.1% NP-40, 10 mM Tris, pH 7.4, 40 U ml<sup>-1</sup> RNase inhibitor, 0.5 mM DTT), and the solution was mixed with rotation at 4 °C for 2 h. For YTHDF1, YTHDF2, YTHDF3, 10  $\mu$ l GST-affinity magnetic beads (Pierce) were used for each sample after being washed four times with 200  $\mu$ l IPP buffer for each wash. For C-YTHDF2, 20  $\mu$ l Dynabeads His-Tag Isolation & Pulldown beads (Invitrogen) were used after being washed four times with 200  $\mu$ l IPP buffer for each wash. The beads were then re-suspended in 50  $\mu$ l IPP buffer. The protein-RNA mixture was combined with GST or His6 beads and kept rotating for another 2 h at 4 °C. The aqueous phase was collected, recovered by ethanol precipitation, dissolved in 15  $\mu$ l water, and saved as the flow-through. The beads were washed four times with 300  $\mu$ l IPP buffer each time. 0.4 ml TRIzol reagent was added to the beads and further purified according to the manufacturer's instructions. The purified fraction was dissolved in 15  $\mu$ l water, and saved as YTHDF-bound. LC-MS/MS was used to measure the level of m<sup>6</sup>A in each sample of input, flow-through and YTHDF-bound.

**LC-MS/MS<sup>7,8</sup>.** 200–300 ng of mRNA was digested by nuclease P1 (2 U) in 25  $\mu$ l of buffer containing 25 mM of NaCl, and 2.5 mM of  $ZnCl_2$  at 37 °C for 2 h, followed by the addition of  $NH_4HCO_3$  (1 M, 3  $\mu$ l) and alkaline phosphatase (0.5 U). After an additional incubation at 37 °C for 2 h, the sample was diluted to 50  $\mu$ l and filtered (0.22  $\mu$ m pore size, 4 mm diameter, Millipore), and 5  $\mu$ l of the solution was injected into LC-MS/MS. Nucleosides were separated by reverse phase ultra-performance liquid chromatography on a C18 column with on-line mass spectrometry detection using an Agilent 6410 QQQ triple-quadrupole LC mass spectrometer in positive electrospray ionization mode. The nucleosides were quantified by using the nucleoside to base ion mass transitions of 282 to 150 ( $m^6A$ ), and 268 to 136 (A). Quantification was performed in comparison with the standard curve obtained from pure nucleoside standards running on the same batch of samples. The ratio of  $m^6A$  to A was calculated based on the calibrated concentrations.

**$m^6A$  profiling.** Total RNA was isolated from HeLa cells with TRIzol reagent. Poly(A)<sup>+</sup> RNA was further enriched from total RNA by using FastTrack MAG Maxi mRNA isolation kit (Invitrogen). In particular, an additional DNase I digestion step was applied to all the samples to avoid DNA contamination. RNA fragmentation,  $m^6A$ -seq, and library preparation were performed according to the previous protocol developed in ref. 14. The experiment was conducted in two biological replicates (Extended Data Table 1).

**RIP-seq.** The procedure was adapted from the previous report<sup>20</sup>. 60 million HeLa cells were collected (three 15-cm plates, after 24 h transfection of Flag-tagged YTHDF2) by cell lifter (Corning Incorporated), pelleted by centrifuge for 5 min at 1,000g and washed once with cold PBS (6 ml). The cell pellet was re-suspended with 2 volumes of lysis buffer (150 mM KCl, 10 mM HEPES pH 7.6, 2 mM EDTA, 0.5% NP-40, 0.5 mM DTT, 1:100 protease inhibitor cocktail, 400 U ml<sup>-1</sup> RNase inhibitor; one plate with ~200  $\mu$ l cell pellet and ~400  $\mu$ l lysis buffer), pipetted up and down several times, and then the mRNP lysate was incubated on ice for 5 min and shock-frozen at -80 °C with liquid nitrogen. The mRNP lysate was thawed on ice and centrifuged at 15,000g for 15 min to clear the lysate. The lysate was further cleared by filtering through a 0.22  $\mu$ m membrane syringe. 50  $\mu$ l cell lysate was saved as input, mixed with 1 ml TRIzol. The anti-Flag M2 magnetic beads (Sigma, 20  $\mu$ l per ml lysate, ~30  $\mu$ l to each sample) was washed with a 600  $\mu$ l NT2 buffer (200 mM NaCl, 50 mM HEPES pH 7.6, 2 mM EDTA, 0.05% NP-40, 0.5 mM DTT, 200 U ml<sup>-1</sup> RNase inhibitor) four times and then re-suspended in 800  $\mu$ l ice-cold NT2 buffer. Cell lysate was mixed with M2 beads; the tube was flicked several times to mix the contents and then rotated continuously at 4 °C for 4 h. The beads were collected, washed eight times with 1 ml ice-cold NT2 buffer. 5 packed beads volumes (~150  $\mu$ l = 30  $\mu$ l  $\times$  5) of elution solution which was 500 ng  $\mu$ l<sup>-1</sup> 3  $\times$  Flag peptide (Sigma) in NT2 buffer were added to each sample, and the mixture was rotated at 4 °C for 2 h to elute. The supernatant was mixed with 1 ml TRIzol and saved as IP. RNA recovered from input was further subjected to mRNA purification by either Poly(A) selection (replicate 1, FastTrack MAG Micro mRNA isolation kit, invitrogen) or rRNA removal (replicate 2, RiboMinus Eukaryote Kit v2, Ambion). Input mRNA and IP with 150–200 ng RNA of each sample were used to generate the library using TruSeq stranded mRNA sample preparation kit (Illumina).

**PAR-CLIP.** We followed the previously reported protocol<sup>32</sup> with the following modifications. Sample preparation: Five 15-cm plates of HeLa cells were seeded at Day 1 18:00. At Day 2 10:00, the HeLa cells were transfected with Flag-tagged YTHDF2 plasmid at 80% confluency. After six hours, the media was changed and 200  $\mu$ M 4SU was added. At Day 3 10:00, the media was aspirated, and the cells were washed once with 5 ml ice-cold PBS for each plate. The plates were kept on ice, and the crosslink was carried out by 0.15 J cm<sup>-2</sup> Ultraviolet light. 2 ml PBS was added and the cells were collected by cell lifter.

Library construction: the final recovered RNA sample was further cleaned by RNA Clean & Concentrator (Zymo Research) before library construction by Tru-seq small RNA sample preparation kit (Illumina).

Mild enzyme digestion<sup>33</sup>. The first round of T1 digest was carried out under 0.2 U  $\mu$ l<sup>-1</sup> for 15 min instead of 1 U  $\mu$ l<sup>-1</sup> for 15 min. The second round of T1 digest was conducted under 10 U  $\mu$ l<sup>-1</sup> for 8 min instead of 50 U  $\mu$ l<sup>-1</sup> for 15 min.

**Ribosome and polysome profiling.** The procedure was adapted from the previous report<sup>22</sup>. Eight 15-cm plates of HeLa cells were prepared for 48 h knockdown (siControl, siYTHDF2, four plates each). Before collection, cycloheximide (CHX) was added to the media at 100  $\mu$ g ml<sup>-1</sup> for 7 min. The media was removed, and the cells were collected by cell lifter with 5 ml cold PBS with CHX (100  $\mu$ g ml<sup>-1</sup>). The cell suspension was spun at 400g for 2 min and the cell pellet was washed once by 5 ml PBS-CHX per plate. 1 ml lysis buffer (10 mM Tris, pH 7.4, 150 mM KCl, 5 mM  $MgCl_2$ , 100  $\mu$ g ml<sup>-1</sup> CHX, 0.5% Triton-X-100, freshly add 1:100 protease inhibitor, 40 U ml<sup>-1</sup> SUPERasin) was added to suspend the cells and then kept on ice for 15 min with occasional pipetting and rotating. After centrifugation at 15,000g for 15 min, the supernatant (~1.2 ml) was collected and absorbance tested at 260 nm (150–200  $A_{260\text{ nm}}$  ml<sup>-1</sup>). To the lysate, 8  $\mu$ l DNase Turbo was added. The lysate was then split by the ratio of 1:4 (Portion I/Portion II). 4  $\mu$ l Super

RNasin was added to Portion I. 40  $\mu$ l MNase buffer and 3  $\mu$ l MNase (6,000 gel units, NEB) was added to Portion II. Both portions were kept at room temperature for 15 min, and then 8  $\mu$ l SUPERasin was added to Portion II to stop the reaction. Portion I was saved and mixed with 1 ml TRIzol to purify input mRNA. Portion II was used for ribosome profiling.

Ribosome profiling: a 10/50% w/v sucrose gradient was prepared in a lysis buffer without Triton-X-100. Portion II was loaded onto the sucrose gradient and centrifuged at 4 °C for 4 h at 27,500 r.p.m. (Beckman, rotor SW28). The sample was then fractionated and analysed by Gradient Station (BioCamp) equipped with ECONO UV monitor (BioRad) and fraction collector (FC203B, Gilson). The fractions corresponding to 80S monosome (not 40S or 60S) were collected, combined, and mixed with an equal volume of TRIzol to purify the RNA. The RNA pellet was dissolved in 30  $\mu$ l water, mixed with 30  $\mu$ l 2  $\times$  TBE-urea loading buffer (Invitrogen), and separated on a 10% TBE-urea gel. A 21-nt and a 42-nt ssRNA oligo were used as size markers, and the gel band between 21 and 42 nt was cut. The gel was passed through a needle hole to break the gel, and 600  $\mu$ l extraction buffer (300 mM NaOAc, pH 5.5, 1 mM EDTA, 0.1 U ml<sup>-1</sup> RNasin) was added. The gel slurry was heated at 65 °C for 10 min with shaking, and then filtered through 1 ml Qiagen filter. RNA was concentrated by ethanol precipitation and finally dissolved in 10  $\mu$ l of RNase-free water.

Input mRNA: the input RNA was first purified by TRIzol and the input mRNA was then separated by PolyAtract. The resulting mRNA was concentrated by ethanol precipitation and dissolved in 10  $\mu$ l of RNase-free water. The mRNA was fragmented by RNA fragmentation kit (Ambion). The reaction was diluted to 20  $\mu$ l and cleaned up by micro Bio-Spin 30 column (cut-off: 20 bp; exchange buffer to Tris).

Library construction: the end structures of the RNA fragments of ribosome profiling and mRNA input were repaired by T4 PNK: (1) 3' de-phosphorylation: RNA (20  $\mu$ l) was mixed with 2.5  $\mu$ l PNK buffer and 1  $\mu$ l T4 PNK, and kept at 37 °C for 1 h; (2) 5'-phosphorylation: to the reaction mixture, 1  $\mu$ l 10 mM ATP and 1  $\mu$ l extra T4 PNK were added, and the mixture was kept at 37 °C for 30 min. The RNA was purified by 500  $\mu$ l TRIzol reagent, and finally dissolved in 10  $\mu$ l water. The library was constructed by Tru-seq small RNA sample preparation kit (Illumina). The sequencing data obtained from ribosome profiling (portion II) were denoted as ribosome-protected fragments and that from RNA input (portion I) as mRNA input. Translation efficiency was defined as the ratio of ribosome-protected fragments and mRNA input, which reflected the relative occupancy of 80S ribosome per mRNA species.

Polysome profiling: sample preparation and sucrose gradient were the same as those of the ribosome profiling procedure except eliminating MNase digestion. The fractions resulting from sucrose gradient were used for western blotting or pooled to isolate total RNA for RT-PCR and mRNA for LC-MS/MS test of  $m^6A/A$  ratio.

**RNA-seq for mRNA lifetime.** Two 10-cm plates of HeLa cells were transfected with YTHDF2 siRNA or control siRNA at 30% confluency. After 6 h, each 10-cm plate was re-seeded into three 6-cm plates, and each plate was controlled to afford the same amount of cells. After 48 h, actinomycin D was added to 5  $\mu$ g ml<sup>-1</sup> at 6 h, 3 h, and 0 h before trypsinization collection. The total RNA was purified by RNeasy kit (Qiagen). Before construction of the library with Tru-seq mRNA sample preparation kit (Illumina), ERCC RNA spike-in control (Ambion) was added to each sample (0.1  $\mu$ l per sample). Two biological replicates were generated: (1) in replicate 1, RNA spike-in control was added proportional to cell numbers; (2) in replicate 2, RNA spike-in control was added proportional to total RNA. Although data obtained from the two sets showed systematic shift, they led to consistent conclusion that YTHDF2 knockdown leads to prolonged lifetime of its RNA targets (Extended Data Fig. 9).

**Data analysis of seq-data.** General pre-processing of reads: All samples were sequenced by illumine Hiseq2000 with single end 100-bp read length. For libraries that generated from small RNA (PAR-CLIP and ribosome profiling), the adapters were trimmed by using FASTX-Toolkit<sup>34</sup>. The deep sequencing data were mapped to Human genome version hg19 by Tophat version 2.0<sup>35</sup> without any gaps and allowed for at most two mismatches. RIP and Ribosome profiling were analysed by DESeq<sup>36</sup> to generate RPKM (reads per kilobase, per million reads). mRNA lifetime data were analysed by Cuffdiff version 2.0<sup>37</sup> to calculate RPKM.

Data analysis for each experiment: (1) for  $m^6A$  profiling, the  $m^6A$ -enriched regions in each  $m^6A$ -immunoprecipitation sample were extracted by using the model-based analysis of ChIP-seq (MACS) peak-calling algorithm<sup>38</sup>, with the corresponding  $m^6A$ -Input sample serving as the input control. For each library, the enriched peaks with  $P < 10^{-5}$  were used for further analysis; (2) for RIP, enrichment fold was calculated as  $\log_2(\text{IP}/\text{input})$ ; (3) PAR-CLIP data were analysed by PARalyzer1.1 with default settings<sup>39</sup>; (4) for ribosome profiling, only genes with RPKM > 1 were used for analysis and the change fold was calculated as  $\log_2(\text{siYTHDF2}/\text{siControl})$ ; (5) for mRNA lifetime profiling: RPKM were converted to attomole by linear-fitting of the RNA spike-in.



The degradation rate of RNA  $k$  was estimated by

$$\log_2 \left( \frac{A_t}{A_0} \right) = -kt$$

where  $t$  is transcription inhibition time (h),  $A_t$  and  $A_0$  represent mRNA quantity (attomole) at time  $t$  and time 0. Two  $k$  values were calculated: time 3 h versus time 0 h, and time 6 h versus time 0 h. The final lifetime was calculated by using the average of  $k_{3h}$  and  $k_{6h}$ .

$$t_{1/2} = \frac{2}{k_{3h} + k_{6h}}$$

**Integrative data analysis and statistics:** PAR-CLIP targets were defined as reproducible gene targets among three biological replicates (3,251). RIP targets (2,528) were genes with  $\log_2(\text{IP}/\text{input}) > 1$ . The overlap of PAR-CLIP and RIP targets were defined as CLIP + IP targets (1,277). And non-targets (3,905) should meet the conditions: (1) complementary set of PAR-CLIP targets; (2) RIP enrichment fold  $< 0$ . For the comparison of PAR-CLIP and m<sup>6</sup>A peaks, at least 1 bp overlap was applied as the criteria of overlap peaks. Two biological replicates were conducted for ribosome profiling and mRNA lifetime profiling, respectively. And genes with sufficient expression level (RPKM  $> 1$ ) were subjected to further analysis. The change fold that used in the main text is the average of the two  $\log_2(\text{siYTHDF2}/\text{siControl})$  values. Nonparametric Mann–Whitney  $U$ -test (Wilcoxon rank-sum test, two sided, significance level = 0.05) was applied in ribosome profiling data analysis as previous reported<sup>22</sup>. For the analysis of cell viability (Extended Data Fig. 8e), RPF of ribosome profiling data were analysed by Cuffdiff version 2.0 for differential expression test, and the genes that differentially expressed ( $P < 0.05$ ) were subjected to Ingenuity Pathway Analysis (IPA, Ingenuity System). RPF was chosen since it may better reflect the translation status of each gene.

**Data accession:** all the raw data and processed files have been deposited in the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>). m<sup>6</sup>A profiling data are accessible under GSE46705 (GSM1135030 and GSM1135031 are input samples whereas GSM1135032 and GSM1135033 are immunoprecipitation samples). All other data are accessible under GSE49339.

**RT–PCR.** Real-time PCR (RT–PCR) was performed to assess the relative abundance of mRNA. All RNA templates used for RT–PCR were pre-treated with on column DNase I in the purification step. The RT–PCR primers were designed to span exon–exon junctions in order to further eliminate the amplification of genomic DNA and unspliced mRNA. When the examined gene had more than one isoform, only exon–exon junctions shared by all isoforms were selected to evaluate the overall expression of that gene. RT–PCR was performed by using Platinum one-step kit (Invitrogen) with 200–400 ng total RNA template or 10–20 ng mRNA template. *HPRT1* was used as an internal control because: (1) *HPRT1* mRNA did not have m<sup>6</sup>A peak from m<sup>6</sup>A profiling data; (2) *HPRT1* mRNA was not bound by YTHDF2 from the PAR-CLIP and RIP sequencing data; (3) *HPRT1* showed relative invariant expression upon YTHDF2 knockdown from the RNA-seq data; (4) *HPRT1* was a house-keeping gene.

**YTHDF2:** TAGCCAACTGCGACACATTC; CACGACCTTGACGTTCTCTT. **SON:** TGACAGATTTGGATAAGGCTCA; GTCCTCCTGACITTTTAGCAA. **CREBBP:** CTCAGCTGTGACCTCATGGA; AGTGTGCTAGTCTCGCACAC. **PLAC2:** AAGCGCTACCACATCAAGGT; CCTCCAACCCAGACTACCTG. **LDLR:** GCTACCCCTCGAGACAGATG; CACTGTCCGAAGCCTGTCT. **HPRT1:** TGACACTGGCAAAACAATGCA; GGTCCCTTTTCCACGACAGCT. **F-luc or F-luc-5BoxB:** CACCTTCGTGACTTCCCATTT; TGACTGAATCGGACACAAGC.

**R-luc:** GTAACGCTGCCTCCAGCTAC; CCAAGCGGTGAGGTACTGT.

A combination of knockdown/overexpression/RIP/RT–PCR experiments was conducted to evaluate the occupancy change of YTHDF2 on its RNA targets after MT-A70 (METTL3) knockdown (Extended Data Fig. 5). Two 15-cm plates of HeLa cells were transfected with siControl or siMETTL3 siRNA. After 10 h, the cells were re-seeded. After 14 h, the cells were further transfected with Flag-tagged YTHDF2 plasmid, and collected after another 24 h (in total 48 h knockdown of METTL3, 24 h over-expression of Flag–YTHDF2). Anti-Flag beads were used to separate YTHDF2-bound portion (IP) from unbound portion (flow-through) as described in the RIP section.

**Fluorescence microscopy.** Fluorescent immunostaining: the protocol of ref. 26 was followed. The cells were grown in an 8-well chamber (Lab-Tek). After treatment indicated in each experiment, the cells were washed once in PBS and then fixed in 4% paraformaldehyde in PBST (PBS with 0.05% Tween-20; prepared by mixing paraformaldehyde with PBST, heat at 60 °C until clear, pH ~7.5) at room temperature for 15 min under rotation. The fixing solution was removed, and –20 °C chilled methanol was immediately added to each chamber and incubated for 10 min at room temperature. The cells were rinsed once in PBS and incubated

with blocking solution (10% FBS with PBST) for 1 h at room temperature under rotation. After that, the blocking solution was replaced with primary antibody (diluted by fold indicated in Antibodies section in blocking solution) and incubated for 1 h at room temperature (or overnight at 4 °C). After being washed 4 times with PBST (300 µl, 5–10 min for each wash), secondary antibody (1:300 dilution in PBST) was added to the mixture and incubated at room temperature for 1 h. After washing 4 times with PBST (300 µl, 5–10 min for each wash), anti-fade reagent (slowfade, Invitrogen) was added to mount the slides.

**FISH in conjugation with fluorescent immunostaining:** Stellaris FISH probe with Quasar 570 was used according to the manufacturer's instructions. After the washing step, the sample preparation proceeded to the blocking step of the previous paragraph in the presence of 40 U ml<sup>–1</sup> of RNase inhibitor. Secondary antibodies were Alexa 488 and Alexa 647 conjugates.

**Image capture and analysis:** the images were captured by Leica SP5 II STED-CW super-resolution laser scanning confocal microscope, analysed by ImageJ. The colocalization was quantified by JAcOP (ImageJ plug-in) and the Pearson coefficients in main text Fig. 3 were gained under Costes' automatic threshold<sup>40</sup>.

**Protein co-immunoprecipitation.** HeLa cells expressing Flag-tagged YTHDF2, N-YTHDF2, C-YTHDF2 or pcDNA3.0 blank vector were collected by cell lifter (three 15-cm plates for each), and pelleted by centrifuge at 400g for 5 min. The cell pellet was resuspended with 2 volumes of lysis buffer (the same as the one used in RIP), and incubated on ice for 10 min. To remove the cell debris, the lysate solution was centrifuged at 15,000 g for 15 min at 4 °C, and the resulting supernatant was passed through a 0.22-µm membrane syringe filter. While 50 µl of cell lysate was saved as Input, the rest was incubated with the anti-Flag M2 magnetic beads (Sigma) in ice-cold NT2 buffer (the same as the one used in RIP) for 4 h at 4 °C. Afterwards, the beads were subject to extensive wash with 8 × 1 ml portions of ice-cold NT2 buffer, followed by incubation with the elution solution containing 3 × Flag peptide (0.5 mg ml<sup>–1</sup> in NT2 buffer, Sigma) at 4 °C for another 2 h. The eluted samples, saved as IP, were analysed by western blotting. For IP samples, each lane was loaded with 2 µg IP portion; and the input lane were loaded with 10 µg Input portion which corresponded to ~1% of overall input.

**Tether assay.** Basic setting: 100 ng reporter plasmid (pmirGlo or pmirGlo-5BoxB) and 500 ng effector plasmid (pcDNA-Flag-λ, pcDNA-Flag-Y2Nλ, or pcDNA-Flag-Y2N) were used to transfect the HeLa cells in each well of six-well plate at 60–80% confluency. After 6 h, each well was re-seeded into 96-well plate (1:20) and 12-well plate (1:2). After 24 h, the cells in 96-well plate were assayed by Dual-Glo Luciferase Assay Systems (Promega). Firefly luciferase (F-luc) activity was normalized by *Renilla* luciferase (R-luc) to evaluate the translation of reporter. And samples in 12-well plate were processed to extract total RNA (DNase I digested), followed by RT–PCR quantification. The amount of F-luc mRNA was also normalized by that of R-luc mRNA.

**RNA immunoprecipitation:** Two 15-cm plates of HeLa cells were transfected with 1 µg pmirGlo-5BoxB reporter and 5 µg pcDNA-Flag-Y2Nλ effector plasmids for each plate. After 24 h, the samples were processed as described in RIP section. The recovered RNA from Input, IP and FT portions were used in poly(A) tail assay.

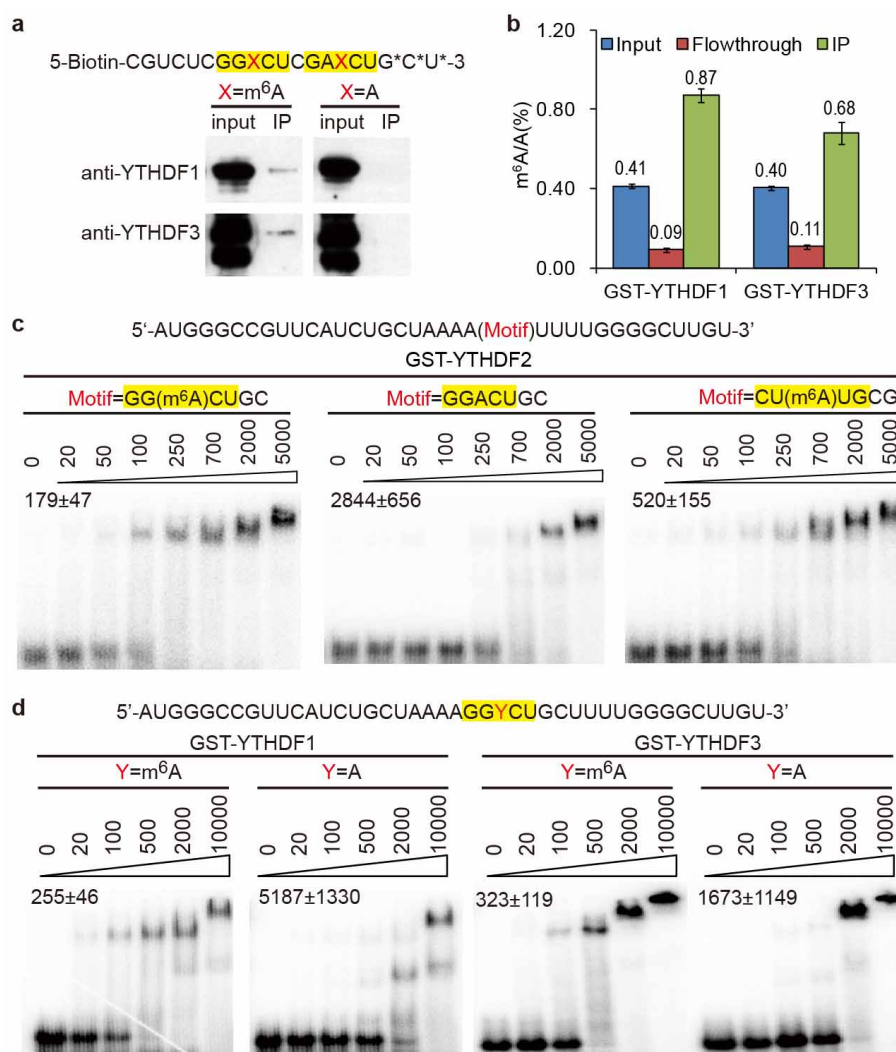
**RNA decay:** 200 ng reporter plasmid (pmirGlo-Ptigh-5BoxB) and 1 µg effector plasmid (pcDNA-Flag-λ, pcDNA-Flag-Y2Nλ, or pcDNA-Flag-Y2N) were used for each 6 cm plate to transfect the HeLa Tet-off cell line (Clontech) in the presence of 400 ng doxycycline (Dox, Clontech). The transcription of F-luc5BoxB was under repression at this stage. After 18 h, the cells in each 6-cm plate were washed twice with PBS, trypsinized, and washed twice with Dox-free media, then split to four equal portions and re-seeded to 12-well plate in Dox-free media. After 4 h pulse transcription of F-luc5BoxB, Dox was added to 400 ng in each well. The first time point ( $t = 0$  h) was taken as after 20 min<sup>41</sup>, then 2 h, 4 h and 6 h. Total RNA extracted from each sample were used for RT–PCR analysis and Poly(A) tail length assay.

**Poly(A) tail length assay.** Poly(A) tail length assay was performed by using Poly(A) Tail-Length Assay kit (Affymetrix) as previously reported<sup>7</sup>. The protocol of the manufacture (Extended Data Fig. 7f–l) was followed, with 30 cycles of two-step PCR at the last step, and then visualized on 10% non-denaturing TBE gel. The forward primer of F-luc-5BoxB is 5'-CCGCTGAGCAATACTAGCA-3', and the gene-specific reverse primer is 5'-TGCAATTGTGTTGTTAACTGTTT-3'. The forward primer of *CREBBP* mRNA is 5'-GTCTTGGGCAATCCAGATGT-3', and the gene-specific reverse primer is 5'-TTTGATCCAAGTAGITTTACCATC-3'.

**Antibodies.** The antibodies used in this study were listed below in the format of name (application; catalogue; supplier; dilution fold): Rabbit anti-YTHDF1 (Western; ab99080; Abcam; 1,000). Rabbit anti-YTHDF3 (Western; ab103328; Abcam; 1,000). Mouse anti-Flag HRP conjugate (Western; A5892; Sigma; 5000). Rabbit anti-MT-A70 (Western; 15073-1-AP; Proteintech Group; 3000). Rabbit anti-FTO (Western; 5325-1; Epitomics; 10,000). Goat anti-GAPDH HRP conjugate (Western; A00192; GeneScript; 15,000). Rabbit anti-DCP2 (Western; Ab28658;

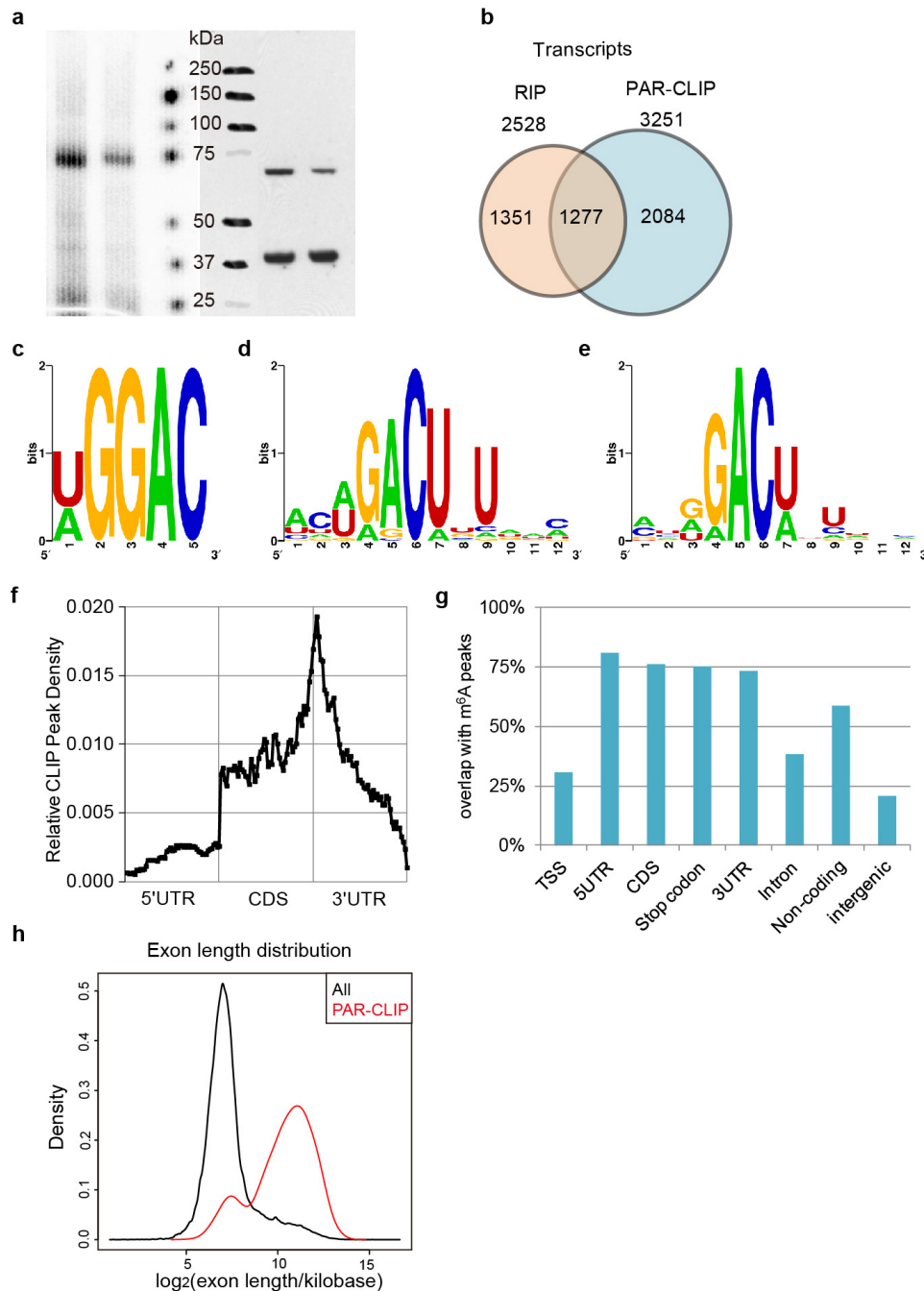


- Abcam; 1,000). Rabbit anti-m<sup>6</sup>A (m<sup>6</sup>A-seq; 202003; Synaptic Systems; 4 µg per seq). Rat anti-Flag (IF; 637304; Biolegend; 300). Mouse anti-DCP1a (IF; WH0055802M6; Sigma; 300). Mouse anti-GW182 (4B6) (IF; ab70522; Abcam; 100). Rabbit anti-DDX6 (IF; a300-461A; Bethyl Lab; 250). Anti-HuR (IF; WH0001994M2; Sigma; 50). Goat anti-eIF3 (N-20) (IF; sc-16377; Santa Cruz Biotech; 100). Mouse anti-CNOT7 (IF; sc-101009; Santa Cruz Biotech; 100). Goat anti-PAN2 (C-20) (IF; sc-82110; Santa Cruz Biotech.; 100). Anti-PARN (IF; ab27778; Abcam; 100). Donkey anti-rat Alexa 488 (IF; A21208; Molecular Probes; 300). Goat anti-rabbit Alexa 647 (IF; A21446; Molecular Probes; 300). Goat anti-mouse Alexa 647 (IF; A21236; Molecular Probes; 300). Donkey anti-goat Alexa 647 (IF; A21447; Molecular Probes; 300).
31. Pillai, R. S., Artus, C. G. & Filipowicz, W. Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA* **10**, 1518–1525 (2004).
  32. Hafner, M. *et al.* PAR-CLIP - a method to identify transcriptome-wide the binding sites of RNA binding proteins. *J. Vis. Exp.* **41**, e2034 (2010).
  33. Kishore, S. *et al.* A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nature Methods* **8**, 559–564 (2011).
  34. Pearson, W. R., Wood, T., Zhang, Z. & Miller, W. Comparison of DNA sequences with protein sequences. *Genomics* **46**, 24–36 (1997).
  35. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
  36. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
  37. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28**, 511–515 (2010).
  38. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
  39. Corcoran, D. L. *et al.* PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. *Genome Biol.* **12**, R79 (2011).
  40. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).
  41. Clement, S. L. & Lykke-Andersen, J. A tethering approach to study proteins that activate mRNA turnover in human cells. *Methods Mol. Biol.* **419**, 121–133 (2008).



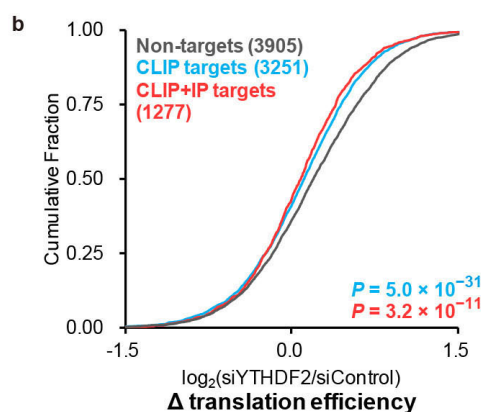
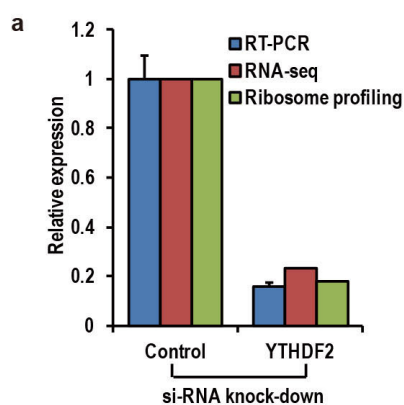
**Extended Data Figure 1 | YTH domain family members are m<sup>6</sup>A-specific RNA binding proteins.** **a**, Western blot showing YTHDF1 and YTHDF3 pulled down with an m<sup>6</sup>A-containing RNA probe. \*Thiol-substituted phosphodiester bonds were used to prevent enzymatic cleavage. **b**, LC-MS/MS showing that m<sup>6</sup>A was enriched in GST-YTHDF1- or GST-YTHDF3-bound mRNA while depleted in the flow-through portion. **c**, **d**, Gel-shift assay

measuring the dissociation constant ( $K_d$ , nM, indicated at the upper left corner of the gel) of GST-tagged YTH domain family proteins (**c**, YTHDF2; **d**, YTHDF1 and YTHDF3) with methylated and unmethylated RNA probes. 4 nmol RNA probe was labelled with <sup>32</sup>P and the protein concentration ranged from 20 nM to 5  $\mu$ M.



**Extended Data Figure 2 | Features and comparisons of YTHDF2 PAR-CLIP data with RIP and  $m^6A$ -seq.** **a**, Left, PAR-CLIP gel image showing  $^{32}$ P-labelled RNA-YTHDF2 complex; right, western blotting of HeLa cell lysate with overexpression of Flag-tagged YTHDF2 (10  $\mu$ g per lane). Upper band was detected by anti-Flag antibody; lower band was detected by anti-GAPDH antibody. **b**, Overlap of transcripts identified by PAR-CLIP and RIP-seq of YTHDF2. **c, d**, YTHDF2 binding motif identified by MEME with top 1,000 scored PAR-CLIP peaks under different motif searching parameters. **c**, With motif length restricted to 5–10 bp,  $P = 1.1 \times 10^{-43}$ , 183 sites were found under this motif. **d**, The motif length was restricted to 5–12 bp. The motif with lowest  $P$  value was shown in main text as Fig. 1c, this motif showed the second lowest  $P$  value,  $P = 5.1 \times 10^{-14}$ , 104 sites were found. **e**, With 7–12 bp,  $P = 7.5 \times 10^{-42}$ ,

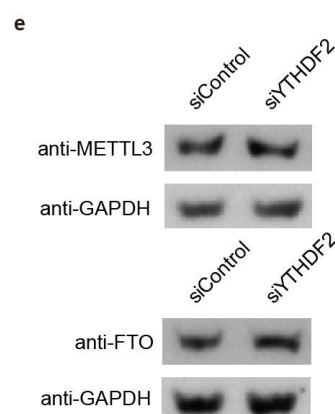
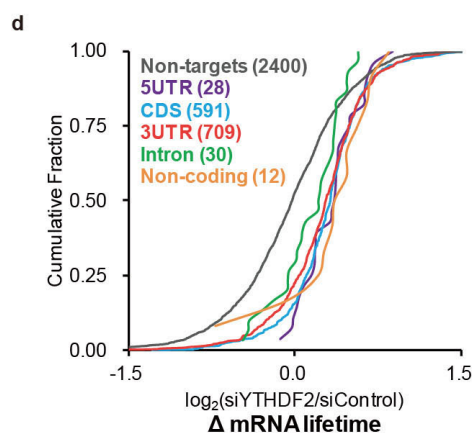
231 sites were found under this motif. **f**, Distribution of PAR-CLIP peaks across the length of mRNA. Each region of 5' UTR, CDS, and 3' UTR were binned into 50 segments, and the percentage of PAR-CLIP peaks that fall within each bin was determined. **g**, Overlap of YTHDF2 PAR-CLIP peaks with  $m^6A$  peaks in different sub-transcript regions. Over 70% PAR-CLIP peaks in 5' UTR, CDS, stop codon, and 3' UTR regions overlap with  $m^6A$  peaks (at least 1-bp overlap). In contrast, only 20%~30% of PAR-CLIP peaks in transcription starting sites (TSS) and intergenic regions coincide with  $m^6A$  peaks. **h**, Enrichment of YTHDF2 PAR-CLIP peaks in long exons. The length distribution of exons that contain YTHDF2 PAR-CLIP peaks (red) shifts to larger size compared with the length distribution of all exons in the human genome (black).



**c**

Sample	Data set	Frequency	Sum of ranks	Mean of ranks	Median	Groups*
mRNA input	NT	3905	15177599	3886.709	-0.154	A
	CLIP	3251	14242532	4380.907	0.038	B
	CLIP+IP	1277	6141829	4809.577	0.125	C
Ribosome-protected fragments	NT	3905	16129368	4130.440	0.063	A
	CLIP	3251	13633808	4193.727	0.086	A
	CLIP+IP	1277	5798784	4540.943	0.147	B
Translation Efficiency	NT	3905	17393203	4454.085	0.189	B
	CLIP	3251	13147547	4044.155	0.110	A
	CLIP+IP	1277	5021210	3932.036	0.084	A
mRNA lifetime	NT	2400	5333012	2222.088	-0.030	A
	CLIP	2504	8159838	3258.721	0.273	B
	CLIP+IP	966	3738535	3870.119	0.325	C
	NT	2400	4815283	2006.368	-0.030	A
	CLIP=1	1034	2730168	2640.395	0.195	B
	CLIP=2~4	1058	3097145	2927.358	0.286	C
	CLIP $\geq$ 5	412	1384464	3360.350	0.383	D
	NT	2400	3868749	1611.979	-0.030	A
	5UTR	28	71550	2555.357	0.359	B
	CDS	591	1432600	2424.027	0.319	B
	3UTR	709	1642005	2315.945	0.273	B
	Intron	30	61929	2064.300	0.146	A
	Non-coding	12	31502	2625.167	0.414	B

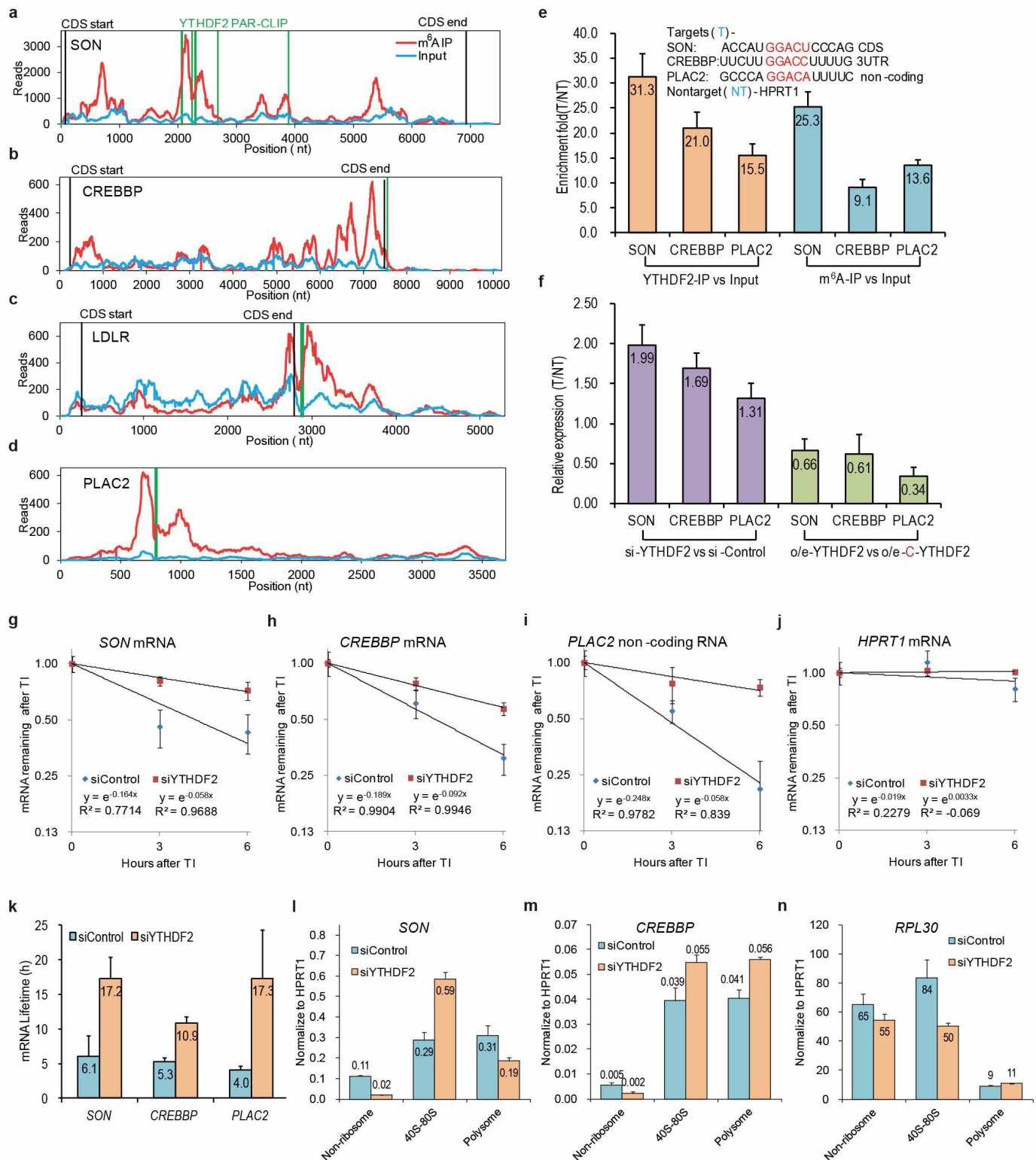
\* group naming follows the sorting of the mean of ranks: A<B<C<D





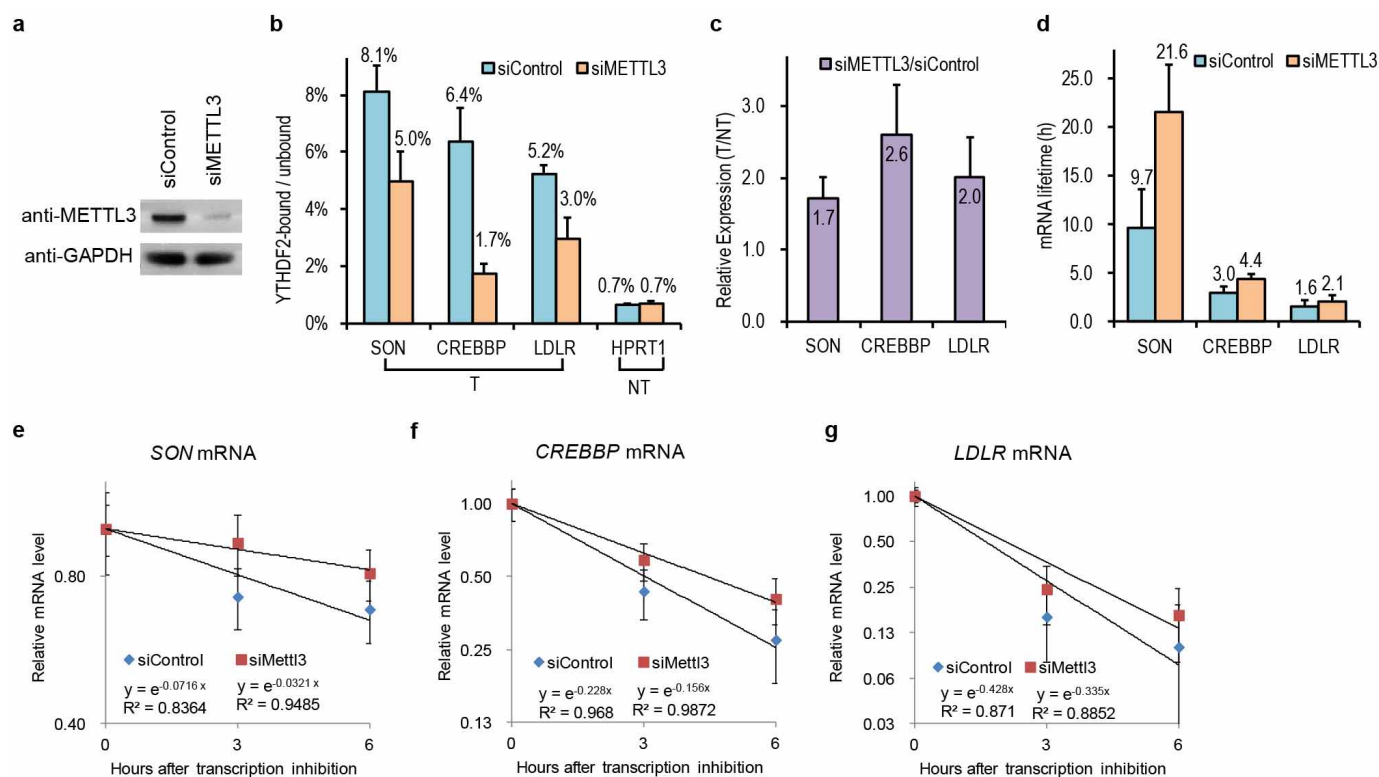
**Extended Data Figure 3 | Effects of YTHDF2 knockdown and summary of the sequencing data.** **a**, The YTHDF2 knockdown efficiency is about 80% as detected by RT-PCR (error bars, mean  $\pm$  s.d.,  $n = 3$ , biological replicates) and RNA-seq. Although at current stage we could not identify a reliable antibody for YTHDF2, ribosome-profiling of YTHDF2 did indicate that the translation level of YTHDF2 decreased by 80% after siRNA knockdown. RT-PCR results were normalized to that of GAPDH as an internal control. RNA-seq and ribosome profiling results were calculated by actual RPKM. **b**, YTHDF2 knockdown led to decreased translation efficiency of its targets due to the accumulation of non-translating mRNA. Translation efficiency is calculated as the ratio of ribosome-protected fragments and mRNA input.  $P$  value was

calculated by using Mann-Whitney  $U$ -test (two-tailed, significance level = 0.05). **c**, Multiple pairwise comparisons (Kruskal-Wallis test) by using the Steel-Dwass-Critchlow-Fligner procedure (two-tailed, significance level = 0.05). **d**, The regional effect of the YTHDF2-binding site is not significant. Cumulative distribution showing mRNA lifetime  $\log_2$ -fold changes ( $\Delta$ ) between si-YTHDF2 and si-control for non-targets and CLIP-IP common targets with major CLIP peak at 5' UTR, CDS, 3' UTR, intron, and non-coding RNA. Except for intron, other regions show similar fold changes (also see Extended Data Fig. 3c). **e**, The m<sup>6</sup>A methyltransferase (MT-A70) and demethylase (FTO) remain unchanged with YTHDF2 knockdown.



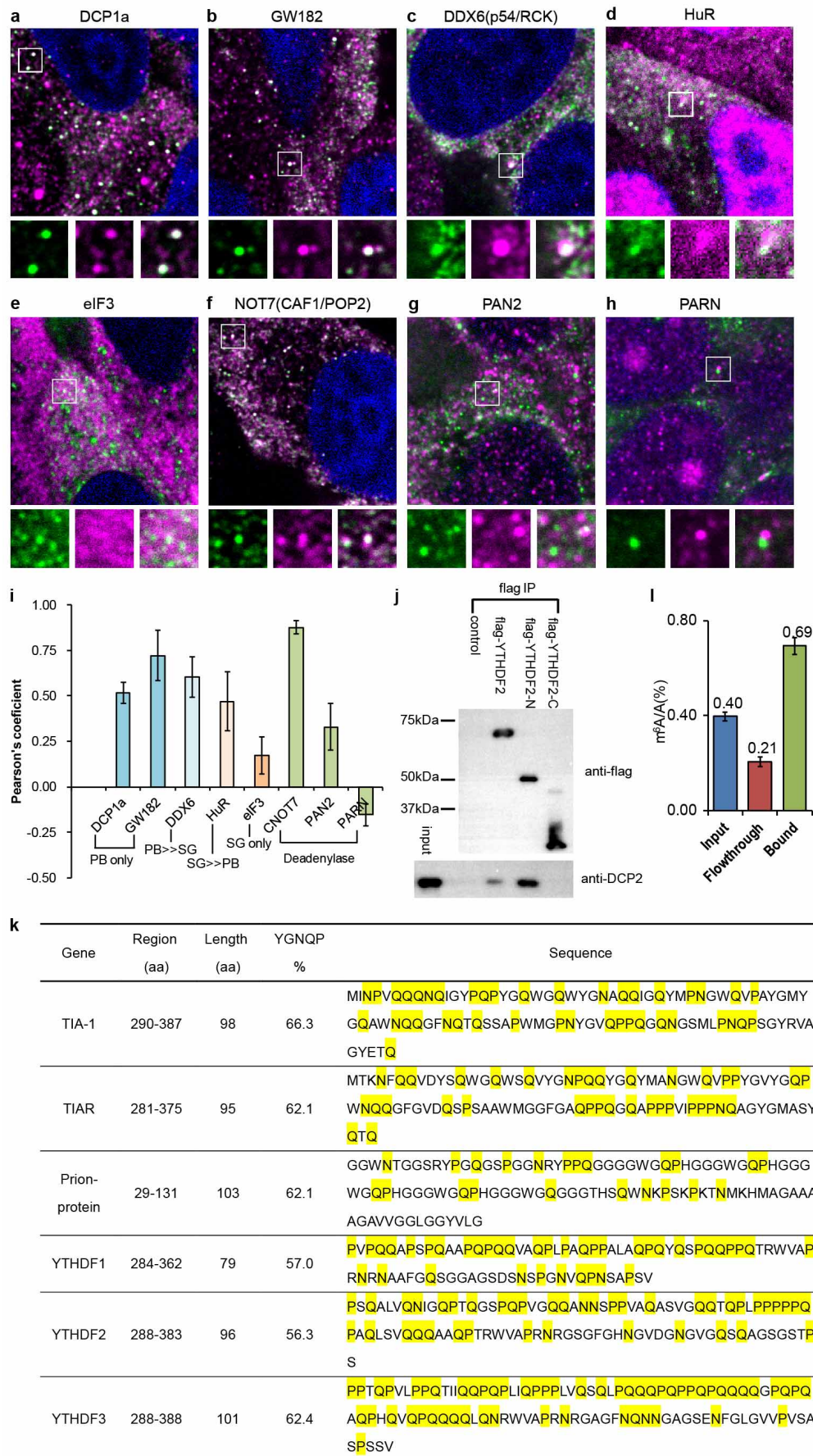
**Extended Data Figure 4 | Validation of representative YTHDF2 RNA targets.** **a–d**, Examples of transcripts harbouring m<sup>6</sup>A peaks and YTHDF2 PAR-CLIP peaks: *SON* (CDS, **a**), *CREBBP* (3' UTR, **b**), *LDLR* (3' UTR, **c**), *PLAC2* (non-coding RNA, **d**). Coverage of m<sup>6</sup>A immunoprecipitation and input fragments are indicated in red and blue, respectively. YTHDF2 PAR-CLIP peaks are highlighted in green. Black lines signify CDS borders. **e–n**, relative RNA level quantified by gene-specific RT-PCR, and error bars shown in these figure panels are mean  $\pm$  s.d.,  $n = 6$  (two biological replicates  $\times$  three technical replicates). **e**, Enrichment fold of *SON*, *CREBBP* mRNA, and *PLAC2* RNA in YTHDF2-RNA coimmunoprecipitation versus RNA-protein input control, and in m<sup>6</sup>A *in vitro* immunoprecipitation versus

mRNA input control. **f**, Relative changes of *SON*, *CREBBP* mRNA, and *PLAC2* RNA in siYTHDF2 sample versus siControl, and overexpression of YTHDF2 versus overexpression of C-YTHDF2. **g–k**, Lifetimes of *SON*, *CREBBP* mRNA and *PLAC2* RNA under siYTHDF2 versus siControl. **l–n**, YTHDF2 knockdown altered the cytoplasmic distribution of its mRNA targets. The *SON* (**l**) and *CREBBP* (**m**) mRNA levels decreased in the non-ribosome mRNP portion but increased in the 40S–80S portion under siYTHDF2 compared to siControl. However, they showed different changes in the polysome portion. *RPL30* (**n**) is not a target of YTHDF2 and did not show an increase in the 40S–80S portion.



**Extended Data Figure 5 | Knockdown of METTL3 (MT-A70) led to decreased binding of YTHDF2 to its targets and increased stability of its target RNAs similar to that of YTHDF2 knockdown.** **a**, Western blotting showing that the knockdown efficiency of siMETTL3 at 48 h was ~80%. **b–g**, Relative RNA level quantified by gene-specific RT-PCR, and error bars shown in these figure panels are mean  $\pm$  s.d.,  $n = 6$  (two biological replicates  $\times$  three technical replicates). **b**, Percentages of YTHDF2 targets (SON, CREBBP, LDLR) in YTHDF2-bound portion versus unbound portion

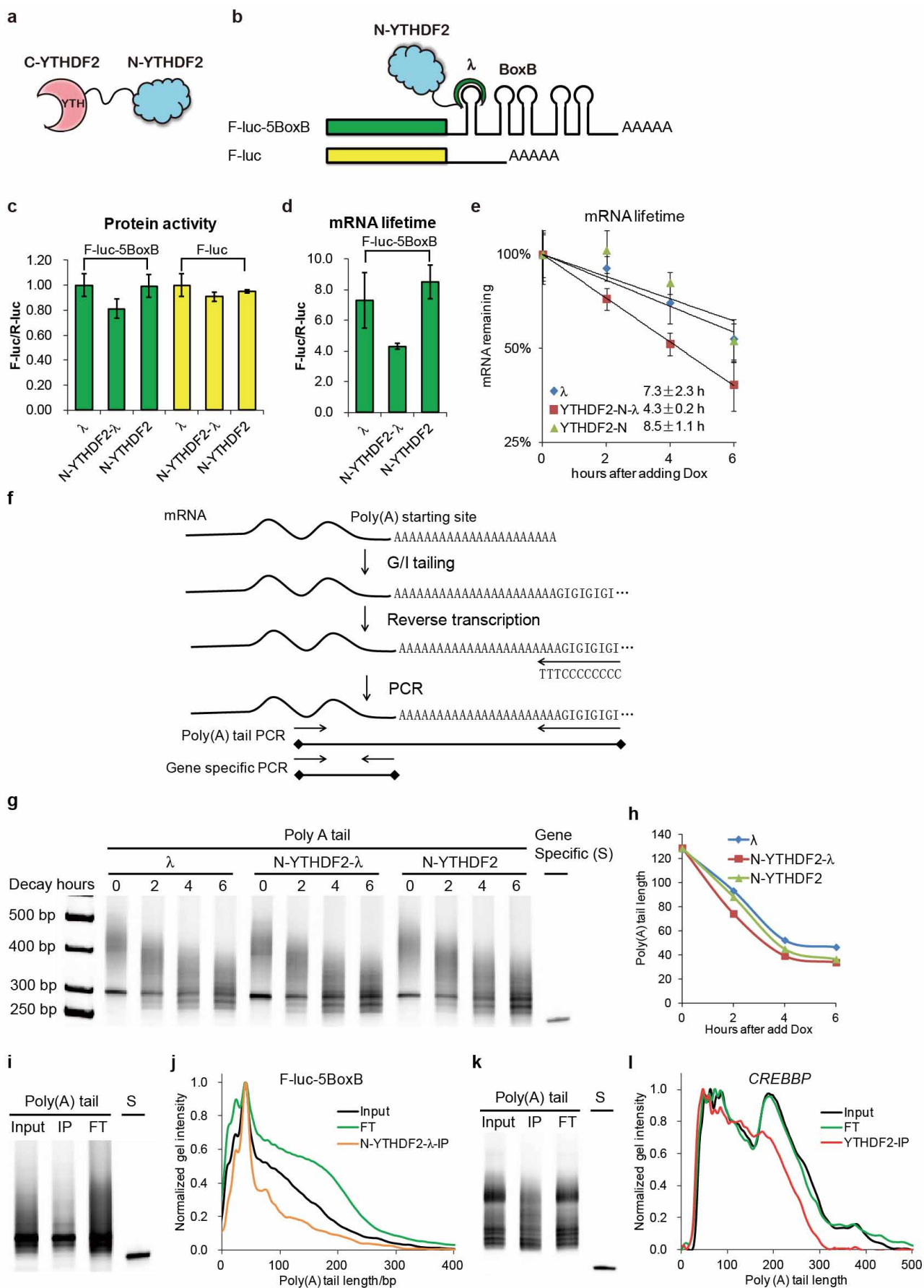
decreased significantly after METTL3 knockdown for 48 h. After 24 h transfection of METTL3 siRNA, HeLa cells were transfected with Flag-tagged YTHDF2, and cells were collected after another 24 h. Anti-Flag beads were used to separate YTHDF2-bound portion (IP) from unbound portion (flow-through). Each transcript was quantified by RT-PCR. **c**, Relative changes of SON, CREBBP and LDLR mRNA in siMETTL3 sample versus siControl. **d–g**, Lifetimes of SON, CREBBP, and LDLR mRNA under siMETTL3 versus siControl.





**Extended Data Figure 6 | Co-localization of YTHDF2 with protein markers of P bodies, stress granules, and deadenylation complexes.** **a–h**, Fluorescence immunostaining of Flag-tagged YTHDF2 (green, anti-Flag, Alexa 488) and other protein markers (DCP1a and GW182 for P bodies and eIF3 for stress granule, DDX6 (also known as RCK/p54) and HuR for both, CNOT7, PAN2, and PARN for deadenylation complex; magenta of Alexa 647 is the colour for the marker, green + magenta = white for the co-localization spot). The scale of the magnified region (white frame) is  $1.8\ \mu\text{m} \times 1.8\ \mu\text{m}$ . **i**, Co-localization between YTHDF2 and different protein markers were characterized by Pearson's coefficient, for each pair,  $n = 5\sim 7$ . YTHDF2 seems to have better co-localization with P bodies than stress granules. It also seems to co-localize best with CNOT7 (also known as CAF1 or POP2) which is a subunit of the

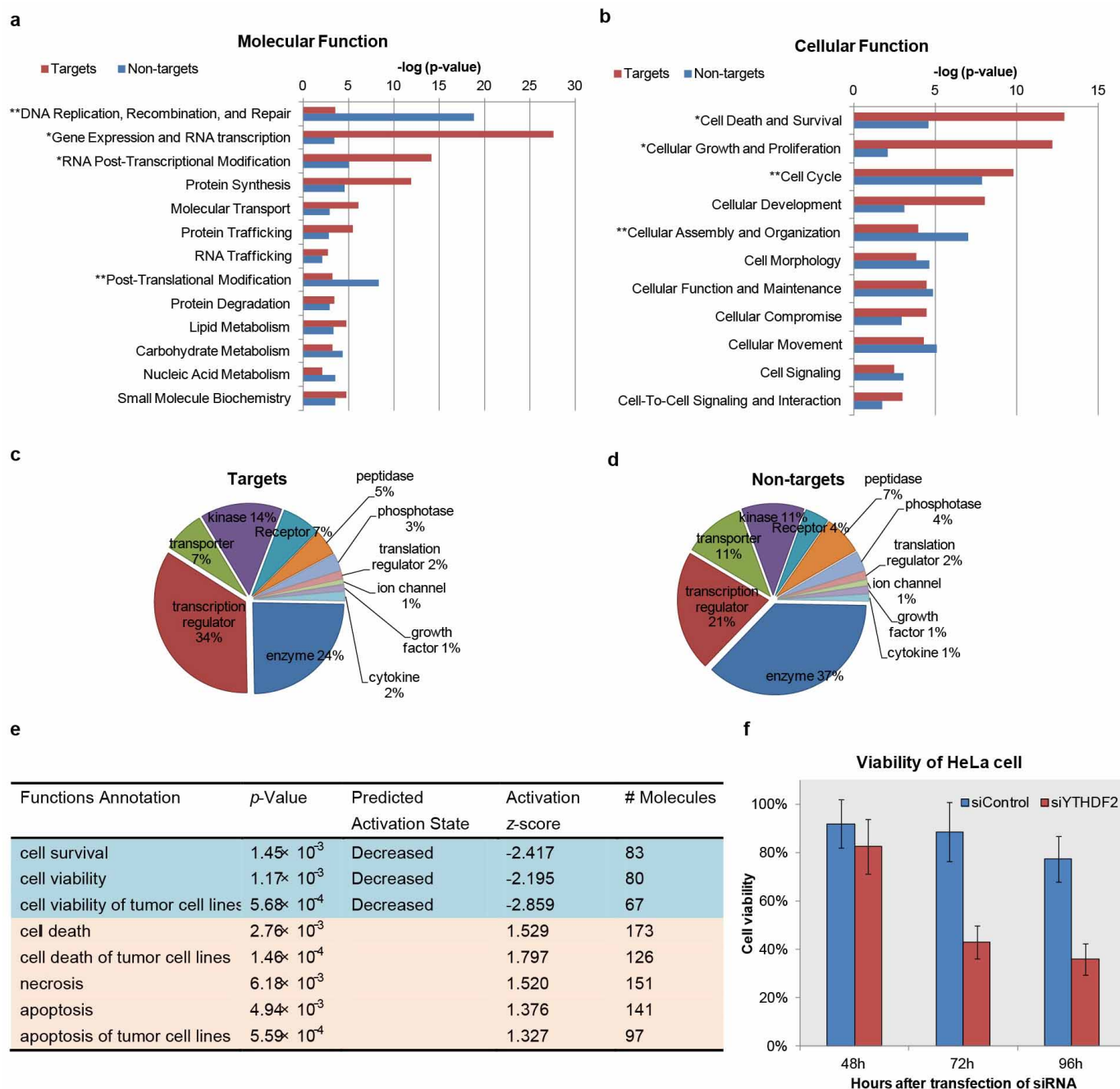
CCR4-NOT deadenylation complex. **j**, Western blotting results showing that immunoprecipitation (IP) of Flag-tagged full length YTHDF2 and N-YTHDF2 (N-terminal domain) also pulled down the P-body marker DCP2, but not with mock control or C-YTHDF2 (the C-terminal domain). For IP samples, each lane was loaded with 2  $\mu\text{g}$  IP portion; and the input lane was loaded with 10  $\mu\text{g}$  input portion which corresponded to  $\sim 1\%$  of overall input). **k**, Comparison of P/Q/N (highlighted) rich regions of YTHDF1-3 with other aggregation-prone proteins. **l**, C-YTHDF2 is capable of selective binding of  $\text{m}^6\text{A}$ -containing RNA. LC-MS/MS showing that  $\text{m}^6\text{A}$ -containing RNA was enriched in the His6-tagged C-YTHDF2-bound mRNA while reduced in the flow-through portion. Error bars shown in the figure are mean  $\pm$  s.d.,  $n = 4$  (two biological replicates  $\times$  two technical replicates).



**Extended Data Figure 7 | Tether assay of the N-terminal domain of****YTHDF2.** **a**, Structural presentation of the two domains of YTHDF2.

**b**, Scheme of the reporter assay: the RNA reporter vector encodes firefly luciferase (F-luc) as the primary reporter and *Renilla* luciferase (R-luc) on the same plasmids acting as transfection control for normalization. Five Box B RNA elements were inserted at the 3' UTR of F-luc as positive tether reporter (noted as F-luc-5BoxB); the effector was a fusion of N-YTHDF2 and  $\lambda$  peptide which recognizes Box B with high affinity. **c**, The F-luc luciferase activity (protein translation) for N-YTHDF2- $\lambda$  was reduced by  $\sim 20\%$  compared to that of N-YTHDF2 and  $\lambda$  controls. Error bars shown in the panel are mean values  $\pm$  s.d. from  $n = 8$  (biological replicates). **d**, **e**, The reporter mRNA lifetime was significantly reduced ( $\sim 40\%$ ) when bound by N-YTHDF2- $\lambda$  as compared to the controls of N-YTHDF2 and  $\lambda$ . Doxycycline (Dox,  $400 \text{ ng } \mu\text{l}^{-1}$ ) was used to inhibit transcription of the reporter. 18 h post transfection of reporter and effectors, Dox was removed to allow a pulse transcription of F-luc-5BoxB for 4 h. Then Dox was added back and the samples were collected at indicated time point. The amounts of F-luc-5BoxB were determined by

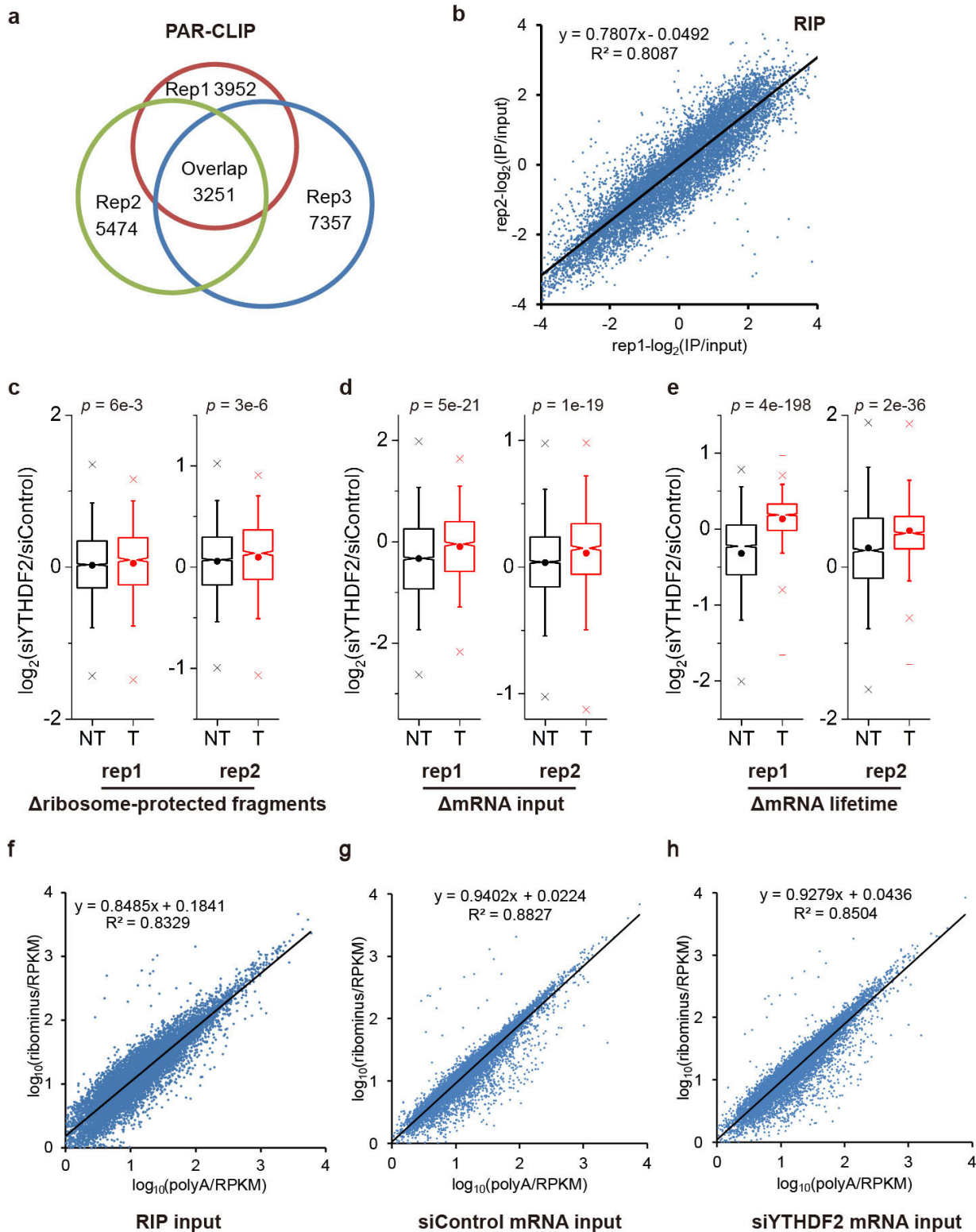
RT-PCR, normalized to R-luc, then for each time series, samples at  $t = 0 \text{ h}$  were set as 100%. Error bars shown in the panel are mean  $\pm$  s.d.,  $n = 6$  (two biological replicates  $\times$  three technical replicates). **f**, Scheme of poly(A) tail length assay. **g**, **h**, Tethering N-YTHDF2 to the reporter mRNA does not significant trigger deadenylation of the reporter. The PCR products of reporter poly(A) tail were visualized in 10% TBE gel stain (**g**) and no significant difference of the deadenylation rate was observed (**h**). **i**–**l**, Shorter poly(A) tail lengths were observed in the YTHDF2-bound fraction for the N-YTHDF2-tethered reporter RNA (**i** and **j**) as well as the native target RNA *CREBBP* (**k** and **l**). Tether reporter F-luc-5BoxB and Flag-tagged YTHDF2-N- $\lambda$  (**i**) or full length Flag-tagged YTHDF2 (**k**) were expressed in HeLa cells, and subjected to immunoprecipitation with anti-Flag beads. RNA recovered from input, IP and flow-through were further processed and the final PCR products for F-luc-5BoxB (**i**) or *CREBBP* (**k**) were visualized in 10% TBE gel. **j** and **l**, each lane were re-plotted against base pair, after *log* fitting of relative gel mobility with base pairs.



**Extended Data Figure 8 | Cellular function of YTHDF2.** **a, b,** The top molecular function of YTHDF2 targets is “Gene Expression and RNA Transcription”, and the top cellular function is “Cell Death and Survival”. Ingenuity Pathway Analysis of function category of YTHDF2 targets and non-targets revealed that the two gene groups are heterogeneous in their functional composition. (\*top two functions for YTHDF2 targets and \*\*top two functions for YTHDF2 non-targets.). **c, d,** Pie charts of molecular types of differentially expressed YTHDF2 targets (**c**) versus non-targets (**d**) upon YTHDF2 knockdown. Differentially expressed genes ( $P$  value  $< 0.05$ ) caused by YTHDF2 knockdown were grouped to YTHDF2 targets (796 gene) and non-targets (1554) based on their presence or absence in YTHDF2 PAR-CLIP binding sites, and subject to Ingenuity Pathway Analysis

(the category “other” was not shown). The results show that the group of YTHDF2 targets is transcription regulators whereas that of non-targets is enzyme, indicating that  $m^6A$  may significantly affect gene expression via tuning mRNA stabilities of transcription factors through YTHDF2. **e, f,** YTHDF2 knockdown led to reduced cell viability. The IPA analysis of ribosome profiling data of YTHDF2 knockdown (48 h) versus control predicts decreased cell viability (**e**). Ribosome profiling data was chosen since it may better reflect the translation status. MTT assay provided experimental evidence of reduced cell viability upon YTHDF2 knockdown.  $P$  values that were calculated from Student’s  $t$ -test were 0.036,  $4.7 \times 10^{-4}$ , and  $9.4 \times 10^{-4}$ , at 48 h, 72 h and 96 h respectively (**f**). Error bars shown in the figure are mean  $\pm$  s.d.,  $n = 10$  (biological replicates).





#### Extended Data Figure 9 | Comparisons of sequencing data with replicates.

**a**, Overlap of three biological replicates (rep1–rep3) for PAR-CLIP. Numbers showing the sum of genes identified in each sample. **b**, Correlation of enrichment fold as  $\log_2(\text{IP/input})$  between two technical RIP replicates. In rep1 the input mRNA was purified by poly(dT) beads, whereas in rep2 the input RNA was processed by rRNA removal. **c–e**, Box plot showing consistent results from two biological replicates that were conducted for ribosome profiling and mRNA lifetime profiling, respectively. For mRNA lifetime profiling, rep1 was normalized by spike-in control that was proportional to cell numbers, whereas rep2 was normalized by spike-in that was proportional to total RNA

concentrations. Despite the technical variations, YTHDF2 knockdown resulted in significant lifetime increase of its targets. (T, 1,277 CLIP+RIP targets; NT, 3,905 non-targets; box, the first and third quartiles; notch, the median; dot in the box: the data average; whisker,  $1.5 \times$  standard deviation; cross, the 1 and 99 percentiles; short line, the maximum and minimum;  $P$  values were calculated by Mann–Whitney  $U$ -test, two-tailed, significant level = 0.05). **f–h**, Correlation of RPKM between technical mRNA input samples prepared by poly(A) selection ( $x$  axis) and by rRNA removal ( $y$  axis), which are comparable to the variations between biological replicates that prepared by the same mRNA selection method.

**Extended Data Table 1 | Summary of the sequencing samples**

Experiment	Sample / replicates	Mapped Reads
PAR-CLIP	rep1	22398274
	rep2	16469620
	rep3	18340002
m <sup>6</sup> A profiling	rep1-input	33037301
	rep1-IP	48299395
	rep2-input	28182497
	rep2-IP	46985896
RIP	IP-rep1	13956595
	IP-rep2	6658433
	input polyA	13995632
	input ribominus	11201129
Ribosome profiling	rep1-SiControl-RPF	6153002
	rep1-SiControl-input-polyA	24693835
	rep1-SiYTHDF2-RPF	4396160
	rep1-SiYTHDF2-input-polyA	23772645
	rep2-SiControl-RPF	10302755
	rep2-SiControl-input-polyA	11276363
	rep2-SiControl-input-ribominus	7336313
	rep2-SiYTHDF2-RPF	9830313
	rep2-SiYTHDF2-input-polyA	9525286
	rep2-SiYTHDF2-input-ribominus	14030008
mRNA lifetime profiling	rep1-SiControl-TI-0h	19703956
	rep1-SiControl-TI-3h	17066177
	rep1-SiControl-TI-6h	25105141
	rep1-SiYTHDF2-TI-0h	23291878
	rep1-SiYTHDF2-TI-3h	18905279
	rep1-SiYTHDF2-TI-6h	27471654
	rep2-SiControl-TI-0h	9709614
	rep2-SiControl-TI-3h	10252359
	rep2-SiControl-TI-6h	13315823
	rep2-SiYTHDF2-TI-0h	9996766
	rep2-SiYTHDF2-TI-3h	11123149
	rep2-SiYTHDF2-TI-6h	7543209

# CAREERS

**BIOMEDICAL RESEARCH** MD–PhD holders focus less on research **p.123**

**WORKFORCE** NIH calls for modelling to mitigate worrying trends **p.123**

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)



COMMUNICATION

## Spontaneous scientists

*Some think that researchers can improve their communication by flexing their improvisation skills.*

BY RACHEL BERNSTEIN

A circle of scientists is gazing skyward, as if watching a ball fly through the air as they play an animated game of catch. But there is no ball — and this game is serious work. It is part of an exercise to help 12 scientists at the University of Connecticut (UConn) Health Center in Farmington to boost their communication skills.

These scientists are engaged in improvisation, a spontaneous, reactive interaction mode more traditionally seen in comedy performances.

Improvisation games and communication both require attention to others and the forging of personal connections, says Raquell Holmes, a cell biologist by training who now spends much of her time running workshops for scientific conferences and research institutions through her company, improvscience, based in Boston, Massachusetts. With that in mind, she has adapted some traditional improvisation exercises, and imaginary catch is one of them.

In this game, participants use eye contact to indicate where the ‘ball’ is being thrown, and use and read body language to communicate its

size and weight so that the recipient can catch it correctly. The skills learned in these games can be directly transferred to scientists’ work pursuits, says Cibele Falkenberg, a computational-biology postdoc at UConn Health Center who has participated in some of the workshops. For instance, she says, “with communication, you have to make sure you have a connection before you pass the message”, which applies to any audience, including co-workers, funding agencies and the public. Convinced of the benefits, improvscience is one of a number of US programmes using improvisation to help hone these skills (see ‘Workshops and events’).

### COMMUNICATIVE COLLABORATION

Researchers sometimes fall short in their communication with each other, despite the importance of collaboration. Holmes thinks that improvisation offers a powerful tool to address this problem — through, for example, the ‘yes, and’ rule. This basic tenet of improvisation dictates that participants must say ‘yes’ to any verbal or physical cues that they receive and build on them, rather than trying to shut down a direction that makes them uncomfortable. The rule is important in a research context, in which a ‘no, but’ stance often dominates — such as when discussing a colleague’s results or critiquing a paper in a journal-club meeting.

From a scientific perspective, this critical approach may be appropriate and necessary. But taken too far, Holmes says, it can create a negative group dynamic and make some people hesitant to share ideas for fear of ridicule. And that, in turn, could slow research progress.

To illustrate this problem, in one of her games Holmes asks participants to get into pairs and work together to plan a party. First, members of each pair can respond to each other’s statements only by starting with ‘no, but’; they then repeat the exercise using the ‘yes, and’ rule. The ‘no, but’ approach made it “very difficult to have a meaningful conversation”, says Max Staller, a systems-biology graduate student at Harvard University in Cambridge, Massachusetts, who has participated in several improvscience workshops. In stark contrast, the ‘yes, and’ rule worked so well in planning the fictitious party that he now applies it to his research.

“I try to consciously think about, is there a way to say ‘yes, and’,” Staller says. “I make a point in journal club of talking about what’s positive about the paper; sometimes we focus too much on the shortcomings, and take for granted the successes.”

Holmes also uses games such as ‘mirrors,’ ►

KUCO/SHUTTERSTOCK

► in which partners have to match each other's body movements as closely as possible, and 'emotional bus', in which participants portray an emotion as they board an imaginary bus, and the rest of the 'passengers' must read the emotion and act it out themselves. The point is to train people to be tuned in and responsive to others around them. "It looks like we're just playing, but what we're practising is how to accept the idea that your colleague is giving you something," explains Falkenberg.

#### AUDIENCE CONNECTION

In some cases, improvisation came to science through the inspired efforts of people who had a penchant for drama. Patricia Ryan Madson, an emeritus lecturer in theatre and performance studies at Stanford University in Palo Alto, California, is no scientist. But when she began teaching improvisational theatre classes to undergraduates in 1991, she found that many of her students were scientists and engineers. It got her thinking about how researchers, too, could benefit from the confidence that practising improvisation inspires and become more comfortable, she says, being "agreeable and helpful and communicative, and making mistakes in a gentle way" that is "not de rigueur in the science world". She found a partner in engineering professor Rolf Faste, and in the late 1990s the two launched an improvisation course that is still offered today.

About 15 years after Madson's initial course, actor and long-time science aficionado Alan Alda came to the idea independently. After hosting a show called *Scientific American Frontiers* from 1993 to 2005, Alda knew he wanted to continue working in science communication. He remembered how his early training in improvisational theatre had boosted his communication skills, and thought it might be valuable for scientists. In January 2008, he tested the idea with a group of engineering students from the University of Southern California in Los Angeles by asking them to explain their research before and after playing improvisation games. "The difference was startling," he recalls. This convinced him that the approach was worth pursuing.

Alda now implements the idea with the Alan Alda Center for Communicating Science at

Stony Brook University in New York, which offers an Improvisation for Scientists course for the university's graduate students and is working to share the curriculum by developing an affiliate network. The first institution to sign up, Dartmouth College in Hanover, New Hampshire, began an improvisation course for scientists in the autumn. The Alan Alda Center also provides workshops for students and faculty members at various conferences and science institutions.

Whereas Holmes's approach focuses primarily on facilitating teamwork, Alda emphasizes



**"With communication, you have to make sure you have a connection before you pass the message."**

Cibele Falkenberg

communicating with non-experts, whether interacting with the media, policy-makers or the public. But the tools are largely the same: encouraging participants to be attuned to others and adapt their behaviour accordingly.

In one Alda Center activity, one member of a pair speaks about a general topic and the other person tries to anticipate what the first will say, so that the two end up saying the same thing at the same time. "It

sounds impossible," says Colin West, a Stony Brook physics graduate who has taken the Alda Center course, "but when you get into a rhythm, you can anticipate a lot based on body language, and the next thing you know you're chanting in unison."

Researchers who are well practised in such an exercise can use their newly honed receptiveness to help them respond to their audiences. "When you're reading the signals," says Alda, "you have instantaneous feedback about whether they're understanding what you're saying. You can adjust what you say because you're accustomed to adjusting." Alda also uses exercises to practise this skill in a scientific context, including one in which a participant presents his or her research to a 'revolving door' of

MARCELO A. KURODA

## IMPROVISATION RESOURCES

### Workshops and events

Improvisation resources are still relatively rare. Both improvscience (improvscience.org) and the Alan Alda Center (centerforcommunicatingscience.org) work with researchers and institutions to set up workshops a few hours to a few days long.

Interested researchers can also attend the 'Improvisation for Scientists: Making a

Human Connection' session on 16 February 2014 in Chicago, Illinois, featuring the Alda Center, improvscience and Katie Watson, who specializes in medical humanities and bioethics at Northwestern University in Evanston, Illinois, and is a faculty member of renowned Chicago-based improvisation group The Second City. **R.B.**





The Alan Alda Center in New York uses improvisation exercises to boost science communication.

audiences. In this game, the instructor assigns the audience an identity, such as a high-school class or an academic tenure committee, without telling the speaker. The audience members behave in keeping with their identity, and the speaker has to interpret their cues to tailor his presentation to them — until the instructor assigns the audience a new identity and the speaker has to adjust seamlessly.

These improvisation exercises helped Philip Fernandes, a graduate student in astrophysics at Dartmouth who recently took the Alda Center course. “I realized that when I give a talk, I memorize it completely because I’m terrified I’m going to go blank, so it turns into this robotic reciting,” he says. Fernandes began watching audiences and reading body language. “If I see that I’m losing them, it’s my job to regain their interest.”

But it is not enough to simply be responsive to the audience. Presenters also need to give the audience something that will help them to forge a connection with the speaker and stay engaged. Alda teaches that the best hook is to include one’s personal perspective, regardless of the audience or venue, so that listeners can relate — although some researchers may not be comfortable with this approach at first because “there’s this idea that emotion doesn’t belong in science”, Fernandes notes. After taking the course and seeing how it improved her communication skills, Dartmouth ecology graduate student Jessica Trout-Haney was convinced that Alda’s approach had merit. “Telling a personal story has made giving talks much more authentic, and fun,” she says. Initially, she had concerns that inserting subjective personal details would signify her science was not objective. “But if we don’t connect with the audience, we’re doing a disservice to the science.” Improvisation helps participants to become more comfortable sharing a side of themselves that they may otherwise hide.

In some exercises, participants must trust the first idea that comes to them — and that thought is instinctive and personal.

Students are videoed giving a short research presentation before and after the course to show them their progress. Krithika Venkataraman, a graduate student in molecular and cellular biology at Stony Brook, says that she saw dramatic results. “I was more enthusiastic, more relaxed when I was up on the stage, and I had a smile on my face, which I didn’t have before,” she says. “I was able to face the audience and be energetic and enthusiastic about my science.”

#### DIVE IN

Trying improvisation for the first time can be nerve-racking, says Christine Urbanowicz, a graduate student in ecology at Dartmouth who took the Alda course. Despite her initial misgivings, she found it rewarding. When the group played imaginary catch during the first meeting, “I kept telling people, ‘don’t pass me the ball, don’t pass me the ball’, even though there wasn’t a ball”, she recalls. But now that she has completed the course, her confidence as a speaker has increased because she knows that she can handle unexpected situations. “Improv allows you to trust yourself enough to know that you’ll be able to figure out where you’re going with your presentation without having it memorized,” she says.

In fact, Urbanowicz has found the experience so helpful that she is spearheading an improvisation group for graduate students at Dartmouth so that she and her classmates can continue to practise and share their skills.

Trout-Haney, for her part, suggests that anyone who is unsure about whether improvisation is useful or right for them should “just take a deep breath and say, ‘yes, and’”. ■

**Rachel Bernstein** is a science writer based in San Francisco.

## PATENTS

### Leaving academia

Scientists who patent their discoveries seem more likely to leave academia than those who do not, says a study published on 5 December (B. Balsmeier and M. Pellens *Econ. Lett.* <http://doi.org/qhvj>; 2013). The authors examined data for 1996–2005 from a survey of 263 academic researchers in Belgium created by the Organisation for Economic Co-operation and Development, and compared responses about career paths with publication and patent records. They found that for each additional patent, up to a total of two, the scientist was approximately a third more likely to leave academia. The authors speculate that patenting reflects an interest in commercializing results, which is better rewarded outside academia.

## BIOMEDICAL RESEARCH

### Physician–scientists

The number of US physician–scientists who conduct biomedical research as their primary profession has declined in the past 30 years, according to a report by the Federation of American Societies for Experimental Biology (FASEB) in Bethesda, Maryland. *Physician Scientists: Assessing the Workforce* finds that from 1982 to 2011, the proportion with medical degrees or MD–PhDs doing research fell from 3.6% to 1.6%. It attributes the decline to factors such as longer training periods and rising debt. But MD–PhD holders have an edge in a tight funding environment because of their skill set, says report co-author Howard Garrison, FASEB’s deputy executive director for policy.

## WORKFORCE

### NIH seeks models

In the wake of its 2012 report on the biomedical workforce (see *Nature* **492**, 167; 2012), the US National Institutes of Health (NIH) is again seeking proposals for computational models of that workforce, with the aim of tracking and mitigating long-term trends that threaten its size and diversity. Such trends include the rising number of biomedical-PhD holders who seek jobs outside academia and the lower numbers of female scientists who stay in academia after PhDs. Michael Sesma, a programme director for the NIH’s National Institute of General Medical Sciences, says that the agency needs models of dynamics, such as how people make career decisions. It will use them to develop grant and other programmes to address problem areas.

# YOU ARE NOT THE FIRST MINION TO DISAPPOINT ME

*It's just so difficult to get the staff these days.*

BY IAN CREASEY

**Y**ou have let me down. Your task was simple! You were supposed to distract Captain Nebulon and lure him into the Shadow Zone, where our phantoms could have toyed with him for years. Now he has slipped your grasp, and he reaches ever closer to the Neutronic Modulator.

You are not the first minion to disappoint me. I'm sure you recall what happened to your predecessor — you oversaw his fate yourself. We always need fresh brains for the Monster Pit.

How could you be so careless? Don't you realize the importance of our scheme? We must obtain the Neutronic Modulator so that we can unlock the wormhole, solve the self-erasing paradox, defeat the Masters of Entropy, and thereby put the governance of the galaxy on a firmer foundation.

I expected so much more from you. Because my previous subordinates had been consistently inept, I employed special measures in your case. I decided that as other people were invariably disappointing, the only person I could rely upon was myself. And so I created you from my own flesh and blood.

Well, flesh. I have very little blood left, after replacing it with athanatic serum. Cancer cells are immortal, so I turned myself into a giant tumour. Eternity is worth a few side effects. My teratogenic outgrowths are mostly non-functional, rarely causing much disturbance outside the nightmares of my enemies. And although I depend upon a constant infusion of fresh serum, I'm careful to maintain back-up systems in case of any interruption to my supply.

I ramble, I know. It is the only way I can hear an intelligent word spoken in the depths of my lair. Certainly I hear little wisdom from you. Even now, you merely cower in front of me, with a vacant expression on your face. I would like to see a little more ambition. When I say "You are not the first minion to disappoint me," I almost wish you would interrupt and say, "But I shall be the last!"

I've told you before how I overthrew my predecessor. Whenever he doubted my loyalty, he used to lecture me on the perils of leadership: "This



spiky titanium throne is not as comfortable as it looks. Being the Hyper-Dimensional Overlord is an endless series of frustrations. Inadequate minions are not the only source of disappointment, I can assure you."

Yet he appreciated my eagerness to concoct new plots rather than merely follow directions. When I deposed him, he admired the efficiency of my coup. I know, because his wraith still haunts this throne and offers me advice.

He always said that you lacked the drive to be a truly cunning henchman. Right now, he is urging me to dispose of you. Perhaps I shall, after I've finished explaining all the ways in which you have disappointed me.

Let's see ... what else is there? You have chosen an incompetent underling, again. Your latest deputy is not as good at slinking undetected as he believes. Slinking is an important skill, one that the younger generation neglects.

You have barely spent any time in the Monster Pit, not even to ride the Chrome Leopards or ingratiate yourself with the Brain Worms. You must feed the Zaptors with your own hands, or you can hardly expect them to snap and snarl when you wish to impress a hostage with their ferocity.

You —

Wait! There is a disturbance in the

dungeons. Someone has breached the perimeter of my back-up vault. Is it Captain Nebulon already? I thought he was busy questing for the Neutronic Modulator. We shall deal with him soon enough, when I have finished reprimanding you. As you seem unable to execute my schemes, perhaps you would like to suggest your own plan for dealing with him?

Speak up! Don't just slump with your jaw gaping. The appropriate time to gape your jaw is when you're terrorizing a victim by displaying your metallic fangs. And you haven't even installed metallic fangs yet. That's yet another sign of your —

*Ouch!* What's happening? The fresh serum is burning me. *Aaargh!* What have you done, you fool?

Oh, I see. You've projected a hologram. Your image has been receiving my rebuke, while you silently sneaked behind me and poisoned my supply of athanatic serum. It's about time! I'm glad to see you finally showing some initiative.

*Uuuuurck. Blarvle. Kkkkss.*

Those are not my last words: I would have uttered something more eloquent. You have succeeded in killing my body, but your deputy has failed to destroy my back-up systems. He blundered around the dungeons without penetrating the inner vault. Inside the necrocomputer, my brain patterns persist. Their output is routed to a panel built into this titanium throne.

Step forward! Switch off the hologram, and reveal yourself. Remove my carcass: its ugliness is offensive. I will permit you to kick it, if you like, but only once. Seat yourself upon the throne, and enjoy your ascent. I look forward to my new advisory role.

You have already learned the frustrations of relying on others to perform crucial assignments. Your first duty as the new Hyper-Dimensional Overlord will be to rebuke your incompetent underling for failing to implement his part of your scheme. Bring him in here. Then repeat after me: "You have let me down. Your task was simple! You are not the first minion to disappoint me ..." ■

**Ian Creasey** lives in Yorkshire, England. His stories have appeared in Asimov's Science Fiction and elsewhere, and his collection *Maps of the Edge* was published in 2011.

JACEY

➔ **NATURE.COM**

Follow Futures:

Twitter @NatureFutures

Facebook go.nature.com/mtoodm